

THE ANNALS OF MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE
OF MATHEMATICAL STATISTICS

Published by the American Statistical Association
and the International Statistical Institute
in cooperation with the International Statistical Congress

Volume 1, Number 1, January 1931

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

Published by the American Statistical Association

and the International Statistical Institute

in cooperation with the International Statistical Congress

STOCHASTIC APPROXIMATION^{1, 2}

By CYRUS DERMAN

Columbia University

1. Introduction. In certain applications, as in bioassay, sensitivity testing, or fatigue trials, the statistician is often interested in estimating a given quantile of a distribution function on the basis of data which is of the zero-one type. For example, suppose $F(x)$ denotes the probability that a metallic test specimen will fracture if subjected to x cycles in a fatigue trial. Then a specimen, when tested in such a way, represents an observation which takes on the value one or zero depending on whether or not it fractures. It is of interest to estimate that number of cycles x such that, for a given α , $F(x) = \alpha$. The techniques of possible use in this connection, such as probit analysis [8] and the "up and down" method of Dixon and Mood [6], depend to a great extent on parametric assumptions concerning the distribution function $F(x)$. Robbins and Monro [13] considered a problem of which the above problem, with or without parametric assumptions, is a special case. Suppose for every real value x , the random variable $Y(x)$, denoting the value of a response to an experiment carried out at a controlled level x , has the unknown distribution function $H(y|x)$ and regression function $M(x) = \int_{-\infty}^{\infty} y dH(y|x)$. Let α be any given real number. Robbins and Monro considered the problem of estimating the root of the equation $M(x) = \alpha$, assuming the existence of a unique root. If $Y(x) = 1$ or 0 with probabilities $F(x)$ and $1 - F(x)$ respectively, where $F(x)$ is a distribution function and $0 \leq \alpha \leq 1$, then $M(x) = F(x)$, and we have the above special case.

The problem of estimating a root of a given regression function has its counterpart in the literature of the more classical mathematics. Newton's method of approximation is, perhaps, the best-known iterative procedure used for such a problem when no random element is present. However, even if $Y(x) = M(x)$ with probability one—i.e., if no randomness exists—Newton's method is not applicable; for Newton's method and other classical procedures depend on knowing the functional form of $M(x)$, whereas, here, such knowledge is not assumed.

Because of the nonparametric nature of the problem, a method of approach not based on the usual curve-fitting techniques, is clearly necessary. As a solution, Robbins and Monro put forward the following iterative scheme. Let $\{a_n\}$ ($n \geq 1$) be a fixed sequence of positive constants such that

$$(1.1) \quad \sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty.$$

Received April 16, 1956.

¹ This research was supported in part by the Office of Scientific Research of the Air Force under Contract AF 18(600)-442, Project R-345-20-7.

² Special Invited Paper read before the Institute of Mathematical Statistics, December 29, 1955, in New York.

The sequence $a_n = 1/n$, for example, satisfies (1.1). Let x_1 , the level of the first experiment, be arbitrary. Succeeding levels are defined recursively by

$$(1.2) \quad x_{n+1} = x_n + a_n(\alpha - y_n),$$

where y_n denotes the response at level x_n —a random variable dependent only on x_n and having the distribution function $H(y | x_n)$. Thus at each stage of experimentation, a new level is chosen, based upon the deviation of the previous response from α and on the number of experiments already performed.

Since the proposal of their scheme, considerable attention has been focused in this direction. Some of this attention has been directed towards establishing conditions under which the Robbins-Monro procedure is reasonable; some has been directed towards treating similar problems with different but analogous schemes; and some has been directed towards providing a more general theory of stochastic approximation. This paper is an exposition of work done along these lines.

2. The Robbins-Monro Process. Let θ be the root of the equation $M(x) = \alpha$. Robbins and Monro [13] proved that x_n defined by (1.2) converges in the mean to θ , i.e., $\lim_{n \rightarrow \infty} E(X_n - \theta)^2 = 0$, in two separate cases. In one case the function $M(x)$ is discontinuous at θ with $|M(x) - \alpha|$ being bounded away from zero for all $x \neq \theta$. (In fact, $M(\theta)$ need not equal α .) In the other case, $M(x)$ is nondecreasing, $M(\theta) = \alpha$, and $M'(\theta) > 0$. In both cases the rather strong condition that $Y(x)$ be bounded with probability one for all x was imposed. However, it should be remarked that for the purpose of estimating a quantile with zero-one data, the condition is not restrictive.

Wolfowitz [16] was the next to take up the problem. He showed that x_n converges to θ in probability under weaker conditions. His most significant improvement was to replace the boundedness condition of Robbins and Monro with the condition that $M(x)$ and $\sigma_x^2 = \int_{-\infty}^{\infty} (y - M(x))^2 dH(y | x)$ be bounded functions of x .

The following conditions, which are weaker than both the Robbins-Monro and Wolfowitz conditions, were assumed by Blum [1].

$$(2.1) \quad |M(x)| \leq c + d|x| \quad \text{for some } c, d \geq 0,$$

$$(2.2) \quad \sigma_x^2 \leq \sigma^2 < \infty \quad \text{for all } x,$$

$$(2.3) \quad M(x) < \alpha \quad (x < \theta), \quad M(x) > \alpha \quad (x > \theta),$$

$$(2.4) \quad \inf_{\delta_1 \leq |x - \theta| \leq \delta_2} |M(x) - \alpha| \quad \text{for every } \delta_1, \delta_2 > 0.$$

Under these assumptions Blum was able to show that $P(\lim_{n \rightarrow \infty} x_n = \theta) = 1$. That this is true, with no stronger assumption than (2.4), is somewhat surprising. For (2.4) allows the possibility that $M(x) \rightarrow \alpha$ as $|x| \rightarrow \infty$, and in such a case one would expect that there might be positive probability of $|x_n|$ converging to ∞ .

While the proofs of Robbins and Monro and of Wolfowitz used arguments

rather special to the process under consideration, Blum's method related to other known results. More specifically, it can be shown using Martingale theory or, more directly, Kolmogorov's inequality that because of (1.1) and (2.2), $\sum_{j=1}^{\infty} a_j(y_j - M(x_j))$ converges with probability one. (See Loève [12], p. 387.) Consequently

$$x_{n+1} - \sum_{j=1}^n a_j(\alpha - M(x_j)) = x_1 - \sum_{j=1}^n a_j(y_j - M(x_j))$$

converges with probability one. Imposing the conditions of (2.1) and (2.3), Blum was able to show that x_n converges with probability one to a random variable W which is finite with probability one. Then (2.4) was enough of a further assumption to allow him to prove that $W = \theta$ with probability one.

Recently, Dvoretzky [7] has shown that under Blum's conditions, x_n also converges in the mean to θ . Dvoretzky's work will be discussed below.

While the results of Blum and Dvoretzky show that under wide conditions the Robbins-Monro process converges to θ both in mean square and with probability one, it is of interest, particularly for statistical purposes, to obtain sharper convergence theorems. To this end, Chung [4] considered two cases. In his first (bounded) case he assumed that

$$(2.5) \quad M(x) = \alpha + \alpha_1(x - \theta) + o(|x - \theta|) \quad (0 < \alpha_1 < \infty),$$

$$(2.6) \quad \inf_{|x-\theta|>\delta} |M(x) - \alpha| = K_0(\delta) > 0 \quad \text{for every } \delta > 0,$$

$$(2.7) \quad P(|Y(x) - \alpha| \leq K_1 < \infty) = 1 \quad \text{for all } x,$$

$$(2.8) \quad \lim_{x \rightarrow \theta} \sigma_x^2 = \sigma_\theta^2 > 0.$$

Under the conditions of (2.5), (2.6), (2.7), and (2.8), Chung showed that if $a_n = 1/(n^{1-\epsilon})$, where $1/[2(1 + K_4)] < \epsilon < \frac{1}{2}$ (K_4 being a constant arising in his analysis), then $n^{(1-\epsilon)/2}(x_n - \theta)$ tends in distribution to the normal distribution with mean 0 and variance $\sigma_\theta^2/(2\alpha_1)$. In his second (quasi-linear) case, he replaced (2.7) with

$$(2.9) \quad K|x - \theta| \leq |M(x) - \alpha| \leq K'|x - \theta| \quad K > 0, K' > \infty,$$

and

$$(2.10) \quad \int_{-\infty}^{\infty} |y - M(x)|^p dH(y|x) \leq K(p) < \infty \quad p = 1, 2, \dots,$$

and showed that if $a_n = c/n$, $c > 1/(2K)$, then $n^{1/2}(x_n - \theta)$ tends in distribution to the normal distribution with mean 0 and variance $(\sigma_\theta^2 c^2)/(2\alpha_1 c^2 - 1)$. Both results were proved by showing the proper convergence of moments. In earlier papers, Kallianpur [10] and Schmetterer [14] and [15] obtained certain bounds for $E(x_n - \theta)^2$. For the most part, however, their results are contained in those of Chung.

The question arises as to whether other limiting distributions might exist.

Chung also showed that all stable laws are possible limiting distributions; and furthermore, no limiting distribution need necessarily exist.

For purposes of application, Chung's results still left something to be desired. Kiefer, who contributed largely to the last section of [4], remarked that, for the quasi-linear case, if $a_n = c/n$, with $c = 1/\alpha_1$, the Robbins-Monro estimate of θ is, under certain regularity conditions, asymptotically minimax if the loss function of an estimate d is $|\theta - d|^r$, $r \geq 0$. That this is true follows with slight modification from results obtained by Wolfowitz [17] on minimax estimation of the mean of a normal distribution with known variance. However, the conditions of the quasi-linear case are not satisfied if $M(x)$ is a distribution function—as is the case in the quantal response problem. Here the bounded case is applicable, but the estimate based on $a_n = 1/n^{1-r}$ has asymptotic efficiency zero. Hodges and Lehman [9], using an idea attributed to Stein, bridged the gap between the quasilinear case and the bounded case, proving that, in the bounded case, $n^{1/2}(x_n - \theta)$ also converges in distribution to the normal distribution with mean 0 and variance $(\sigma_\theta^2 c^2)/(2\alpha_1 c^2 - 1)$ if $a_n = c/n$, $c > 1/(2K'')$, where $K'' = \inf_{|x-\theta| < A} |M(x) - \alpha|/|x - \theta|$ (A being any positive number such that $K'' > 0$). It is not known whether the moments of $n^{1/2}(x_n - \theta)$ converge to the moments of the limiting distribution. Their method was to show, using Blum's probability one convergence theorem, that the asymptotic distribution of x_n depends only on those values of $M(x)$ defined in the neighborhood of $x = \theta$. Within any finite interval, a function $M(x)$ satisfying the conditions of the bounded case will also satisfy the conditions of the quasi-linear case, so that as far as the asymptotic distributions are concerned, the two cases are the same.

Coming back to the quasi-linear case, it has been remarked that for $a_n = 1/(\alpha_1 n)$ and loss function $|\theta - d|^r$, the Robbins-Monro procedure is asymptotically minimax over all possible procedures. Dvoretzky [7] has shown that if it is known that $|x_1 - \theta| \leq C \leq [(2\sigma^2)/K(K' - K)]^{1/2}$, where σ^2 is defined by (2.2), then the choice of $a_n = (KC^2)/(\sigma^2 + nK^2C^2)$ yields estimates such that

$$(2.11) \quad E(x_n - \theta)^2 \leq \frac{\sigma^2 C^2}{\sigma^2 + (n-1)K^2 C^2} \quad n \geq 1,$$

and if any other coefficients are used, there exists an x_1 and a function $M(x)$ satisfying the quasi-linear conditions such that (2.11) does not hold. Except for the case $K = \alpha_1$, Dvoretzky's coefficients lead to estimates having asymptotic variance larger than that obtained by letting $a_n = 1/(\alpha_1 n)$. This loss in asymptotic efficiency is, of course, the price paid for small-sample optimality.

Lehmann and Hodges raised the questions as to how much agreement exists between asymptotic and small-sample theory and how c , if one uses the coefficients $a_n = c/n$, is to be chosen if $\alpha_1 = M'(\theta)$ is unknown. Since the behavior of the variance of the estimate is unknown for $c < 1/(2K)$, they remarked that one would be tempted, if an a priori guess is to be made, to overestimate c . They would also overestimate c on the grounds that $(\sigma_\theta^2 c^2)/(2\alpha_1 c - 1)$, the asymptotic variance of $n^{1/2}(x_n - \theta)$, increases more slowly for increasing $c >$

$1/\alpha_1$ than for decreasing $c < 1/\alpha_1$. In order to gain more insight concerning the proper choice of n and c , they considered the special case of a linear $M(x)$ and constant σ_x^2 . Here it is possible to compute exact variances for all n and to study the effect of varying c on the exact variance. They found, for this special case, rather close agreement between asymptotic and small-sample theory for $n = 20$ and $c > 1/\alpha_1$. However, for $c < 1/\alpha_1$, it appears that the rate of approach of the small-sample variance to the asymptotic variance is much slower, and thus the danger due to underestimating c is, perhaps, not as great as the asymptotic theory suggests.

3. The Kiefer-Wolfowitz Process. Kiefer and Wolfowitz [11] considered the problem of estimating the value of $x = \theta$ such that $M(x)$ is maximum, assuming the existence of a unique maximum. They suggested the following scheme. Let $\{a_n\}$ and $\{c_n\}$ be sequences of positive numbers such that

$$(3.1) \quad c_n \rightarrow 0, \quad \sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n c_n < \infty, \quad \sum_{n=1}^{\infty} a_n^2 c_n^{-2} < \infty.$$

For example, $a_n = 1/n$, $c_n = 1/n^{\frac{1}{2}}$ are such sequences. Let x_1 be arbitrary. Then define recursively

$$(3.2) \quad x_{n+1} = x_n + a_n c_n^{-1} (y_{2n} - y_{2n-1}),$$

where y_{2n-1} and y_{2n} are independent random variables with respective distributions $H(y | x_n - c_n)$ and $H(y | x_n + c_n)$. They proved under certain regularity conditions that x_n converges in probability to θ . Using the same method as with the Robbins-Monro process, Blum [1], under weaker conditions, showed that the process converges with probability one. It also turns out that the condition $\sum a_n c_n < \infty$ is unnecessary. The conditions of Blum and of Kiefer and Wolfowitz are of such a nature that functions like $M(x) = e^{-x^2}$, $-x^2$ are ruled out. Derman [5] considered functions which might be called "quasi-parabolic," analogous to Chung's quasi-linear functions—i.e., functions whose difference quotients lie between two straight lines with positive slopes. Functions like $-x^2$ satisfy these conditions. It was shown in such cases that x_n converges in the mean to θ and in more restrictive cases, where $M(x)$ is locally parabolic at θ , there is, with proper normalization, convergence to the normal distribution. Burkholder [3] also obtained results pertaining to asymptotic normality.

The weakest set of conditions for convergence of the Kiefer-Wolfowitz process which allow both e^{-x^2} and $-x^2$ were given by both Burkholder and Dvoretzky. Burkholder proved probability one convergence and Dvoretzky proved both convergence with probability one and in the mean square. These conditions in Dvoretzky's form are as follows:

$$(3.3) \quad |M(x+1) - M(x)| < A|x| + B < \infty \quad \text{for all } x \text{ and some } A, B,$$

$$(3.4) \quad \sup_{1/k < x - \theta < k} \bar{D}M(x) < 0, \quad \inf_{1/k < 0 - x < k} \underline{D}M(x) > 0 \quad \text{for } k = 1, 2, \dots,$$

$$(3.5) \quad \sigma_x^2 < \sigma^2 < \infty,$$

where $\bar{D}M(x)$ and $\underline{D}(x)$ denote the upper and lower derivates of $M(x)$ at x and σ_x^2 is as in (2.2).

An undesirable feature of the Kiefer-Wolfowitz process is that two observations must be taken at each stage of experimentation. This of course raises the question of whether there exists a procedure having desirable convergence properties which requires only one observation at each stage.³ No such procedure has yet been suggested. However, the approach taken by Dvoretzky appears as if it might allow results in this direction. In the cases that Derman considered, it also turns out that the Kiefer-Wolfowitz procedure yields estimates which have zero asymptotic efficiency—i.e., if x_n is any estimate based on one set of coefficients $\{a_n\}$ and $\{c_n\}$, there exists another estimate x'_n based on coefficients $\{a'_n\}$ and $\{c'_n\}$ such that $\lim_{n \rightarrow \infty} E(x'_n - \theta)^2 / E(x_n - \theta)^2 = 0$. Thus a better procedure seems desirable from this point of view. It is of some interest to note that for cases where $M(x)$ is symmetric about θ , the Robbins-Monro procedure may be used. More explicitly, let ϵ be a small positive number and let $\bar{M}(x) = M(x + \epsilon) - M(x - \epsilon)$ and $y'_n = y_{2n} - y_{2n-1}$, where y_{2n} and y_{2n-1} are observations at $(x + \epsilon)$ and $(x - \epsilon)$ respectively. Then, since $\bar{M}(x)$ is a monotone function of x , and θ is the value of x such that $\bar{M}(x) = \alpha = 0$, the Robbins-Monro procedure $x_{n+1} = x_n - a_n y'_n$ is applicable. Burkholder has pursued this idea further into cases where $M(x)$ is not necessarily symmetric. In such cases, if x_n converges, it converges to a constant which will, in general, differ from θ .

4. Other Procedures. Blum [2] has considered multidimensional analogues to the above problems. Let $Y(x)$ be a k -dimensional random vector with joint distribution function $H(y | x)$, where x is also a k -dimensional vector, and let $M(x)$ denote the expectation of $Y(x)$, where by this we mean that the i th component of $M(x)$ is the expectation of the i th component of $Y(x)$. Conditions were found to ensure that, for a given vector α , a multidimensional version of the Robbins-Monro procedure converges with probability 1 to a vector $x = \theta$, where $M(\theta) = \alpha$. Suppose $Y(x)$ is a random variable which is dependent on x , a k -dimensional vector, and has expectation $M(x)$, a function of x assumed to have a unique maximum. Conditions were also found such that a generalization of the Kiefer-Wolfowitz procedure ($k + 1$ observations at each stage) would yield estimates converging to the vector $x = \theta$, where $M(\theta)$ is maximum. Martingale theory was employed in the convergence proofs.

Returning to one dimension, Burkholder [3] investigated a process slightly more general than either the Robbins-Monro or the Kiefer-Wolfowitz process. Burkholder's process is of the form

$$(4.1) \quad x_{n+1} = x_n + a_n z_n,$$

where $\{a_n\}$ is a sequence of positive numbers and z_n is a random variable with distribution function $H_n(z | x_n)$ —i.e. the distribution functions and therefore the regression functions depend on n . For example, $M_n(x) = (1/c_n)(M(x + c_n) - M(x - c_n))$ in the Kiefer-Wolfowitz procedure is a function of n . Using methods

³ As a matter of fact, this was the original problem concerning the maximum of a regression function posed by H. Robbins.

of Blum, Chung, and Lehmann and Hodges, Burkholder was able to prove various convergence theorems concerning his process. His results carry over to situations where x_n converges to a nonconstant random variable—this occurring when there is no uniqueness of, say, the root of $M(x) = \alpha$ or the maximum of $M(x)$. As a special case of his process, he exhibited a procedure which converges to the point of inflection of a function; e.g., estimating the maximum of a density function with zero-one data is a particular application of such a procedure. Another application permitted by his more general procedure is that of estimating, by taking additional observations at each stage, unknown constants of interest such as α_1 and σ_θ^2 , arising in Section 2.

5. A more general approach to stochastic approximation. A more general approach taken by Dvoretzky [7], viewing a stochastic approximation procedure as a convergent deterministic procedure with a superimposed random element, has proved to be enlightening. For example, suppose $T_n(x_1, \dots, x_n)$ is any transformation of an n -dimensional Euclidean space \mathcal{E}_n into the real numbers such that for some $x = \theta$,

$$(5.1) \quad |T_n(x_1, \dots, x_n) - \theta| \leq F_n |x_n - \theta| \quad \text{for all } (x_1, \dots, x_n) \in \mathcal{E}_n,$$

where F_n is a sequence of positive numbers satisfying

$$(5.2) \quad \prod_{n=1}^{\infty} F_n = 0.$$

Suppose

$$(5.3) \quad x_{n+1} = T_n(x_1, \dots, x_n) + Y_n,$$

where Y_n ($n = 1, \dots$) are random variables such that $E(Y_n | x_1, \dots, x_n) = 0$ and $\sum_{n=1}^{\infty} \sigma_n^2 = \sum_{n=1}^{\infty} EY_n^2 < \infty$. Then, putting $V_n^2 = E(x_n - \theta)^2$ and using (5.1), we have

$$(5.4) \quad V_{n+1}^2 \leq F_n^2 V_n^2 + \sigma_n^2.$$

Let $b_{n-v} = \prod_{i=n-v+1}^n F_i^2$. On iterating (5.4) we get

$$(5.5) \quad V_{n+1}^2 \leq \sum_{i=1}^{n-1} \sigma_i^2 b_{n-i} + \sigma_n^2 + V_1^2 b_{n-1}.$$

For every fixed v , $b_{n-v} \rightarrow 0$ as $n \rightarrow \infty$ by (5.2); and since $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$, it follows, assuming $V_1^2 < \infty$, that the right side of (5.5), and consequently the left, tends to 0 as $n \rightarrow \infty$. Thus any stochastic approximation procedure given by (5.3), with T_n satisfying (5.1) and (5.2) and x_1 chosen such that $V_1^2 < \infty$, yields an estimate which converges in the mean to θ . For the Robbins-Monro scheme, $T_n = x_n + a_n(\alpha - M(x_n))$ and $Y_n = a_n(M(x_n) - y_n)$. Under certain restrictive conditions, (5.1) and (5.2) hold. However, in order that this approach be more generally applicable, it is necessary to weaken condition (5.1). For example, such a weakening is that for sequences of non-negative real numbers α_n , β_n , and γ_n , satisfying $\lim_{n \rightarrow \infty} \alpha_n = 0$, $\sum_{n=1}^{\infty} \beta_n < \infty$, $\sum_{n=1}^{\infty} \gamma_n = \infty$,

$$(5.6) \quad |T_n(x_1, \dots, x_n) - \theta| \leq \max(\alpha_n, (1 + \beta_n) |x_n - \theta| - \gamma_n).$$

A further weakening permits $\alpha_n, \beta_n, \gamma_n$ to depend on x_1, \dots, x_n . Under such conditions Dvoretzky was able to prove that the process (5.3) converges to θ in the mean and with probability one. These conditions are weak enough to apply to the Robbins-Monro and Kiefer-Wolfowitz processes, yielding results mentioned above.

One might expect, then, that whenever a convergent deterministic iteration procedure converges, its stochastic counterpart given by (5.3) will also converge. Dvoretzky constructed a counterexample to show that this is not the case. Thus the conditions $E(Y_n | x_1, \dots, x_n) = 0$ and $\sum_{n=1}^{\infty} E Y_n^2 < \infty$ are not strong enough to allow conditions like (5.6) to be removed.

A further advantage of this general approach is that the convergence theorems hold, with appropriate changes, if x is an element of a normed linear space. Such generality is useful since, in many applications, x will not be a one-dimensional variable. For example, the multidimensional cases treated by Blum can be considered from this point of view.

REFERENCES

- [1] J. R. BLUM, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382-386.
- [2] J. R. BLUM, "Multidimensional stochastic approximation methods," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 737-744.
- [3] D. L. BURKHOLDER, "On a certain class of stochastic approximation processes," Doctoral Thesis, University of North Carolina, 1955.
- [4] K. L. CHUNG, "On a stochastic approximation method," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 463-483.
- [5] C. DERMAN, "An application of Chung's Lemma to the Kiefer-Wolfowitz stochastic approximation procedure," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 532-536.
- [6] W. J. DIXON AND A. M. MOOD, "A method for obtaining and analyzing sensitivity data," *J. Amer. Stat. Assn.*, Vol. 43 (1948), pp. 109-126.
- [7] A. DVORETZKY, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, to be published.
- [8] D. J. FINNEY, *Probit analysis*, Cambridge University Press, 1947.
- [9] J. L. HODGES, JR., AND E. L. LEHMANN, "The Robbins-Monro process in the bounded case," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, to be published.
- [10] G. KALLIANPUR, "A note on the Robbins-Monro stochastic approximation method," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 386-388.
- [11] J. KIEFER AND J. WOLFOWITZ, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462-466.
- [12] U. LÖÈVE, *Probability theory*, D. Van Nostrand Co., Inc., 1955.
- [13] H. ROBBINS AND S. MONRO, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400-407.
- [14] L. SCHMETTERER, "Bemerkungen zum verfahren der stochastischen iteration," *Osterreichisches Ingenieur-Archiv*, Vol. 7 (1953), pp. 111-117.
- [15] L. SCHMETTERER, "Zum sequentialverfahren von Robbins und Monro," *Monatshefte für Mathematik*, Vol. 58 (1954), pp. 33-37.
- [16] J. WOLFOWITZ, "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Stat.* Vol. 23 (1952), pp. 457-461.
- [17] J. WOLFOWITZ, "Minimax estimation of the mean of a normal distribution with known variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 218-230.

CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE PRESENCE OF INFINITELY MANY INCIDENTAL PARAMETERS

BY J. KIEFER¹ AND J. WOLFOWITZ²

Cornell University

Summary. It is shown that, under usual regularity conditions, the maximum likelihood estimator of a structural parameter is strongly consistent, when the (infinitely many) incidental parameters are independently distributed chance variables with a common unknown distribution function. The latter is also consistently estimated although it is not assumed to belong to a parametric class. Application is made to several problems, in particular to the problem of estimating a straight line with both variables subject to error, which thus after all has a maximum likelihood solution.

1. Introduction. Let $\{X_{ij}\}$, $i = 1, \dots, n$, $j = 1, \dots, k$, be chance variables such that the frequency function of X_{i1}, \dots, X_{ik} is $f(x | \theta, \alpha_i)$ when θ and α_i are given, and thus depends upon the unknown (to the statistician) parameters θ and α_i . The parameter θ , upon which all the distributions depend, is called "structural"; the parameters $\{\alpha_i\}$ are called "incidental". Throughout this paper we shall assume that the X_{ij} are independently distributed when $\theta, \alpha_1, \dots, \alpha_n$, are given, and shall consider the problem of consistently estimating θ (as $n \rightarrow \infty$). The chance variables $\{X_{ij}\}$ and the parameters θ and $\{\alpha_i\}$ may be vectors. However, for simplicity of exposition we shall throughout this paper, except in Example 2, assume that they are scalars. Obvious changes will suffice to treat the vector case.

Very many interesting problems are subsumed under the above formulation. Among these is the following:

$$(1.1) \quad f(x | \theta, \alpha_i) = (2\pi\theta)^{-k/2} \exp \left\{ - \frac{\sum_j (x_{ij} - \alpha_i)^2}{2\theta} \right\}.$$

Suppose now that the $\{\alpha_i\}$ are considered as unknown constants and we form in the usual manner the likelihood function

$$(1.2) \quad (2\pi\theta)^{-kn/2} \exp \left\{ - \frac{1}{2\theta} \sum_{i,j} (X_{ij} - \alpha_i)^2 \right\}$$

corresponding to (1.1). Then the maximum likelihood (m.l.) estimator of θ is

$$(1.3) \quad \frac{\sum_{i,j} (X_{ij} - \bar{X}_i)^2}{kn}$$

Received September 28, 1955.

¹ Research sponsored by the Office of Naval Research.

² The research of this author was supported in part by the United States Air Force under Contract AF 18(600)-685, monitored by the Office of Scientific Research.

with $\bar{X}_i = k^{-1} \sum_j X_{ij}$, and is obviously not consistent. This example is due to Neyman and Scott [1], who used it to prove that the m.l. estimator³ need not be consistent when there are infinitely many incidental parameters (constants). The latter authors, to whom the names "structural" and "incidental" are due, seem to have been the first to formulate the general problem. Special forms of the problem, like Example 2 below, had been studied for a long time (e.g., Wald [2] and the literature cited there).

The general fact that, when the $\{\alpha_i\}$ are unknown constants, the m.l. estimator of θ need not be consistent, is certainly basically connected with the fact that, since there are only a constant number of observations which involve a particular α_i , it is in general impossible to estimate the $\{\alpha_i\}$ consistently. Now there are many meaningful and practical statistical problems where the $\{\alpha_i\}$ are not arbitrary constants but independently and identically distributed chance variables with distribution function (df) G_0 (unknown to the statistician). The question then arises whether the m.l. method, which does not always yield a consistent estimator when there are infinitely many incidental constants, and does yield consistent estimators in the classical parametric case where there are no incidental parameters, will give a consistent estimator in this case, where the $\{\alpha_i\}$ are independent chance variables with the common df G_0 . This note is devoted to this question.

The answer is affirmative. Not only is the m.l. estimator of θ strongly consistent (i.e., converges to θ with probability one) under reasonable regularity conditions, but also the m.l. estimator of G_0 converges to G_0 at every point of continuity of the latter, with probability one (w.p.1). This is the more striking when one recalls that G_0 does not belong to a parametric class, i.e., a set of df's indexed by a finite number of parameters. (If G_0 were a member of such a given class, the problem would fall completely in the domain of classical maximum likelihood.) The interest of the present authors was originally in estimating θ . That G can also be estimated by the m.l. method is a felicitous by-product of our investigation. A heuristic explanation of the present result may be this: A sequence of chance variables is more "regular" than an arbitrary sequence of numbers. In the present procedure one does not attempt to determine the particular values of the chance variables $\{\alpha_i\}$, but only their distribution function; thus, we seek the m.l. estimator of the "parameter" $\gamma = (\theta, G)$ based on a sequence of independent random variables whose common distribution function is indexed by γ .

In sections 3, 4, and 5, the results are applied to three problems which seem to be of interest per se. Among these is the problem of fitting a straight line with both variables subject to normal error. This problem has a very long history and has been the subject of many investigations (see, for example [2], [7], [4], and the literature cited there); it seems interesting that it can, after all, be treated by the m.l. method. The verification of the regularity assumptions or the formulation of not too onerous conditions for them to be verified is sometimes not entirely ob-

³ Throughout this paper, for the sake of brevity, we use the term "estimator" to mean "sequence of estimators for $n = 1, 2, \dots$ ".

vicious, and the verification of these assumptions (in the form used in Section 2) constitutes the main difficulty of the paper. As is explained in detail below, the fact that these assumptions imply the general consistency result of Section 2 follows from a modification of the proof of [5]. Professor Herbert Robbins has kindly called our attention to his abstract in *Ann. Math. Stat.*, vol. 21 (1950), p. 314, Abstract 35, which states that the m.l. estimator of G is consistent. Since nothing further has appeared on this subject, the intended restrictions under which the statement is true, and the intended method of proof, are unknown to the present authors. This seems to be the second instance in the literature where the m.l. estimator has been used to estimate an entire df which is not assumed to belong to a class depending only on a finite number of real parameters. The first instance of the employment of such an estimator is the classical estimation of a df by its empiric df (shown to be asymptotically optimal in [3]), which is its m.l. estimator (see the paragraph preceding the lemma in Section 2). The only other instance of the estimation of a df in the nonparametric case seems to be that of the estimation of identifiable df's in stochastic structures such as those of the present paper by means of the minimum distance method [4].⁴ (The latter requires regularity conditions weaker than those of the present paper. Compare, for example, [4] with Example 2 below; see also Example 3a.)

In connection with these examples, and also in Section 6, we give some examples which illustrate the fact that the classical m.l. estimator may not be consistent, even in parametric examples which lack the pathological discontinuity sometimes present in hitherto published examples.

Section 6 also contains the statement of a simple device which can be used in the classical parametric case as well as in the case studied in this paper, to prove consistency of the m.l. estimator in some cases where the assumptions used in published proofs of consistency are not satisfied.

The proof in Section 2 is a modification of Wald's [5], and its fundamental ideas are to be found in [5]; for this reason some of its details will be omitted. Wald states in his paper that his method applies more generally when his Assumption 9 is fulfilled. However, this assumption is not fulfilled in our problem *ab initio* and some technical modifications have to be made. One obstacle to extending Wald's proof to our problem is in establishing an analogue of (16) in [5]; one "neighborhood of infinity" does not always seem to suffice. Also some changes in the assumptions are made necessary by the nature of our problem. The results of the present paper can be extended in the usual manner to abstract spaces, but we forego this. It should also be remarked that in [6] Wald studied the present problem of estimating a structural parameter.

The attitude towards the $\{\alpha_i\}$, i.e., whether they are to be regarded as unknown constants or identically and independently distributed chance variables or something else, seems to vary with the author and sometimes even within the

⁴ A paper entitled "The minimum distance method," which gives the details and proofs of the results announced in [4], is scheduled for publication in a forthcoming issue of these *Annals*.

publications of the same author. For example, Wald [2], in his treatment of the problem of fitting a straight line mentioned above, considers the $\{\alpha_i\}$ as unknown constants; and Neyman and Scott, in their general formulation of the problem given in [1] and described at the beginning of the present section, also consider the $\{\alpha_i\}$ as unknown constants. On the other hand, Neyman in his treatment [7] of the straight line problem treats the α_i as independently and identically distributed chance variables. Also Neyman and Scott [8] criticize Wald's solution [2] on the ground that the conditions he postulates on the sequence of constants $\{\alpha_i\}$ are such that they are unlikely to be satisfied when the $\{\alpha_i\}$ are independently and identically distributed chance variables. Our own point of view and perhaps also that of the other writers cited, is that one need not insist on any one formulation to the exclusion of all others. There are certainly reasonable statistical problems where the $\{\alpha_i\}$ may be looked upon as independently and identically distributed chance variables, and consequently the problem of the present paper is statistically meaningful and interesting. This is also the attitude implicit in [4] and [9].

2. Proof of consistency. As we have stated earlier, the essential idea of the proof comes from [5]. A compactification device has to be employed because the space Γ defined below may not be compact.

We postulate that the following assumptions are fulfilled (see also the paragraph preceding the lemma at the end of this section):

ASSUMPTION 1: $f(x | \theta, \alpha)$ is a density with respect to a σ -finite measure μ on a Euclidean space of which x is the generic point. (This is also Wald's Assumption 1.)

Let Ω be the space of possible values of θ , and let A be the space of values which α_i can take. (Both Ω and A are measurable subsets of Euclidean spaces, f is jointly measurable in x and α for each θ , and we hereafter denote by $\theta_i^{(s)}$ ($1 \leq s \leq r$) the components of a point θ_i in Ω and by $|\alpha|$ the Euclidean distance from the origin of a point $\alpha \in A$; τ will denote Lebesgue measure on A .) Let $\Gamma = \{G\}$ be a given space of (cumulative) distributions of α_i . Let θ_0, G_0 be, respectively, the "true" value of the parameter θ and the "true" distribution of α_i . It is assumed that $\theta_0 \in \Omega$ and $G_0 \in \Gamma$. Let $\gamma = (\theta, G)$ be the generic point in $\Omega \times \Gamma$. We define

$$(2.1) \quad f(x | \gamma) = \int_A f(x | \theta, z) dG(z)$$

and $\gamma_0 = (\theta_0, G_0)$. In the space $\Omega \times \Gamma$ we define the metric

$$(2.2) \quad \begin{aligned} \delta(\gamma_1, \gamma_2) &= \delta([\theta_1, G_1], [\theta_2, G_2]) \\ &= \sum_{s=1}^r |\arctan \theta_1^{(s)} - \arctan \theta_2^{(s)}| \\ &\quad + \int_A |G_1(z) - G_2(z)| e^{-|z|} d\tau(z). \end{aligned}$$

Let $\bar{\Omega} \times \bar{\Gamma}$ be the completed space of $\Omega \times \Gamma$ (the space together with the limits of its Cauchy sequences in the sense of the metric (2.2)). Then $\bar{\Omega} \times \bar{\Gamma}$ is compact.

ASSUMPTION 2 (Continuity Assumption): It is possible to extend the definition of $f(x | \gamma)$ so that the range of γ will be $\bar{\Omega} \times \bar{\Gamma}$ and so that, for any $\{\gamma_i\}$ and γ^* in $\bar{\Omega} \times \bar{\Gamma}$, $\gamma_i \rightarrow \gamma^*$ implies

$$(2.3) \quad f(x | \gamma_i) \rightarrow f(x | \gamma^*),$$

except perhaps on a set of x whose probability is 0 according to the probability density $f(x | \gamma_0)$. (The exceptional x -set may depend on γ^* ; $f(x | \gamma^*)$ need not be a probability density function.) (This assumption corresponds to Wald's continuity Assumptions 3 and 5.)

ASSUMPTION 3: For any γ in $\bar{\Omega} \times \bar{\Gamma}$ and any $\rho > 0$, $w(x | \gamma, \rho)$ is a measurable function of x , where

$$w(x | \gamma, \rho) = \sup f(x | \gamma'),$$

the supremum being taken over all γ' in $\bar{\Omega} \times \bar{\Gamma}$ for which $\delta(\gamma, \gamma') < \rho$. (This assumption is made for the reasons given by Wald. See his remarks following Assumption 8 in [5].)

ASSUMPTION 4 (Identifiability Assumption): If γ_1 in $\bar{\Omega} \times \bar{\Gamma}$ is different from γ_0 , then, for at least one y ,

$$(2.4) \quad \int_{-\infty}^y f(x | \gamma_1) d\mu \neq \int_{-\infty}^y f(x | \gamma_0) d\mu,$$

the integral being over those x all of whose components are \leq the corresponding components of y . (This is the same as Wald's Assumption 4.)

Let X be a chance variable with density $f(x | \gamma_0)$. The operator E will always denote expectation under γ_0 ; γ_0 will always, of course, be a member of $\Omega \times \Gamma$.

ASSUMPTION 5 (Integrability Assumption): For any γ in $\bar{\Omega} \times \bar{\Gamma}$ we have, as $\rho \downarrow 0$,

$$(2.5) \quad \lim E \left[\log \frac{w(X | \gamma, \rho)}{f(X | \gamma_0)} \right]^+ < \infty.$$

(This assumption is implied by assumptions corresponding to Wald's Assumptions 2 and 6.)

For any γ in $\bar{\Omega} \times \bar{\Gamma}$ other than γ_0 , define $v = \log f(X, \gamma) - \log f(X, \gamma_0)$. We begin the proof of consistency by showing that

$$(2.6) \quad Ev < 0.$$

First, if γ is in $\Omega \times \Gamma$, $Ee^v \leq 1$. Hence from (2.3) and Fatou's lemma it follows that, for any γ in $\bar{\Omega} \times \bar{\Gamma}$,

$$(2.7) \quad Ev \leq Ee^v \leq 1.$$

If v is $-\infty$ with probability one according to $f(x | \gamma_0)$, then (2.6) is obvious. Suppose therefore that $v > -\infty$ with positive probability according to

$f(x | \gamma_0)$. Then, by Jensen's inequality and (2.7),

$$(2.8) \quad Ev \leq \log Ee^v \leq 0,$$

and the first equality sign can hold only if v is a constant c with probability one according to $f(x | \gamma_0)$. If the first equality sign does not hold (2.6) follows at once. Consider, therefore, the constant c . If $c < 0$ then (2.6) holds. If $c > 0$ then (2.8) is violated. We cannot have $c = 0$ because of Assumption 4. This proves (2.6).

Now, as $\rho \downarrow 0$, for $\gamma' \neq \gamma_0$,

$$(2.9) \quad \lim E \left[\log \frac{w(X | \gamma, \rho)}{f(X | \gamma_0)} \right]^+ = E \left[\log \frac{f(X | \gamma)}{f(X | \gamma_0)} \right]^+$$

by (2.3), (2.5), and Lebesgue's dominated convergence theorem. Also,

$$(2.10) \quad \lim E \left[\log \frac{w(X | \gamma, \rho)}{f(X | \gamma_0)} \right]^- = E \left[\log \frac{f(X | \gamma)}{f(X | \gamma_0)} \right]^-,$$

since the integrand of the left member decreases monotonically to the integrand of the right member. Hence, as $\rho \rightarrow 0$,

$$(2.11) \quad \lim E \left[\log \frac{w(X | \gamma, \rho)}{f(X | \gamma_0)} \right] = E \log \frac{f(X | \gamma)}{f(X | \gamma_0)} < 0$$

by (2.6). Just as in [5] (see also [12]) it may then be shown that, for any positive ρ , there exists an $h(\rho)$, $0 < h(\rho) < 1$, such that the probability is one that, for all n sufficiently large,

$$(2.12) \quad \sup \left\{ \frac{\prod_{i=1}^n f(X_i | \gamma)}{\prod_{i=1}^n f(X_i | \gamma_0)} \right\} < h^n,$$

the supremum being taken over all γ in $\bar{\Omega} \times \bar{\Gamma}$ for which $\delta(\gamma, \gamma_0) > \rho$, and where X_1, X_2, \dots are independent chance variables with the common density $f(x | \gamma_0)$.

Let $L(x_1, \dots, x_n | \gamma) = \prod_{i=1}^n f(x_i | \gamma)$. A *modified m.l. estimator* is defined to be a sequence of μ -measurable functions $\{\hat{\gamma}_n\}$ such that

$$L(x_1, \dots, x_n | \hat{\gamma}_n(x_1, \dots, x_n)) \geq c \sup_{\gamma} L(x_1, \dots, x_n | \gamma)$$

for almost all $(\mu) x_1, \dots, x_n$ for each n , where c is a positive number (the supremum is over $\Omega \times \Gamma$); for $c = 1$, this of course defines an m.l. estimator. (We shall not be concerned in this paper with conditions which ensure the existence of such measurable functions, although reasonable conditions are not difficult to formulate.) We also define a *neighborhood m.l. estimator* to be a sequence of μ -measurable functions $\{\gamma_n^*\}$ such that there exists a sequence of positive numbers ϵ_n with $\lim_{n \rightarrow \infty} \epsilon_n = 0$ for which $\sup_{\gamma \in \Pi_n} L(x_1, \dots, x_n | \gamma) = \sup_{\gamma} L(x_1, \dots, x_n | \gamma)$ for almost all $(\mu) x_1, \dots, x_n$, where Π_n is the set of all γ in $\Omega \times \Gamma$ for which

$\delta(\gamma, \gamma_n^*(x_1, \dots, x_n)) < \epsilon_n$. (Thus, neighborhood m.l. estimators exist in some cases where m.l. and modified m.l. estimators do not; this will be useful in making clear certain examples below where the lack of consistency is not merely due, as it might seem, to the fact that no strict m.l. or modified m.l. estimator exists.)

The above result (2.12) implies the strong convergence of m.l., modified m.l., and neighborhood m.l. estimators (in the respective cases where they exist). The component of the estimator which estimates G_0 converges to it at all its points of continuity w.p.1.

We remark that the above proof actually demonstrates consistency if, in the definition of m.l. estimator (or its variants), the supremum is taken over $\bar{\Omega} \times \bar{\Gamma}$ instead of over $\Omega \times \Gamma$ or, in fact, over any subset of $\bar{\Omega} \times \bar{\Gamma}$ containing γ_0 . This last fact implies that if consistency is verified in an example where $\Omega = \Omega_1$, $\Gamma = \Gamma_1$, then it automatically holds in the example where $\Omega = \Omega_2$, $\Gamma = \Gamma_2$, whenever $\Omega_2 \subset \Omega_1$ and $\Gamma_2 \subset \Gamma_1$. In particular, this remark applies to the examples of Sections 3, 4, and 5.

We remark that Assumption 1 is not really essential in the above proof. Let P_γ denote the probability measure of X when γ is the true parameter value, and let $d(x, \gamma, \gamma_0) = r(x, \gamma, \gamma_0)/[1 - r(x, \gamma, \gamma_0)]$, where $r(x, \gamma, \gamma_0)$ denotes a Radon-Nikodym derivative of P_γ with respect to $P_\gamma + P_{\gamma_0}$ at the point x . If, for each $\gamma_0 \in \Omega \times \Gamma$, Assumptions 2 and 3 are satisfied when $f(x | \gamma)$ is replaced by $d(x, \gamma, \gamma_0)$, if (2.4) is replaced by the condition that $d(x, \gamma, \gamma_0) = 1$ does not hold on a set of probability one under γ_0 for any γ , and if $f(x | \gamma)/f(x | \gamma_0)$ is replaced by $d(x, \gamma, \gamma_0)$ (with a similar replacement for $w(x | \gamma, \gamma_0)$) in Assumption 5 and in the argument of the section, then (2.12) (with the replacement noted above) will still hold. An m.l. estimator $\hat{\gamma}$ is now defined to be one for which $\sup_\gamma \prod_{i=1}^n d(X_i, \gamma, \hat{\gamma}) = 1$ (with an analogous definition of modified and neighborhood m.l. estimator). We have not stated our assumptions and result (2.12) in this more general setting above because the stated form of the assumptions will suffice in most applications and will be easier to verify than assumptions stated in terms of $d(x, \gamma, \gamma_0)$ (which must be verified for each γ_0). As an example of the use of the more general result just cited, consider the problem of estimating the df F of a sequence of independent identically distributed discrete random variables, it being assumed that the true probability measure P_F (corresponding to the df F) satisfies

$$\sum_x P_F(x) \log P_F(x) > -\infty,$$

the sum being over all points x for which $P_F(x) > 0$. Then the assumptions are easily seen to be satisfied, and we may conclude that the sample df, which is the m.l. estimator, is a consistent estimator of F , a well-known result which does not usually seem to be considered as an example of the consistency of the m.l. estimator. (Of course, even if no restrictions of discreteness or logarithmic summability are placed on P_F , the sample df is still consistent and, as pointed

out in the introduction, this is the m.l. estimator. However, Assumption 5 is not satisfied in this case.)

Before proceeding to our examples in subsequent sections, we prove a simple lemma which will be useful later in verifying Assumption 5.

LEMMA. If $f(z_1, \dots, z_k)$ is a bounded probability density function with respect to Lebesgue measure μ on Euclidean k -space R^k , and if

$$(2.13) \quad \int_{|z_i| > 1} (\log |z_i|) f d\mu < \infty \quad (1 \leq i \leq k),$$

then

$$(2.14) \quad - \int_{R^k} f \log f d\mu < \infty.$$

PROOF: If we prove that (2.13) implies (2.14) when f is replaced by cf in these equations, where $c > 0$, then the lemma is clearly proved. Thus, since f was assumed bounded, we may hereafter assume $f \leq (2e)^{-1}$. (The new f need not have integral unity.) Let

$$(2.15) \quad g(z_1, \dots, z_k) = f(z_1, \dots, z_k) + \prod_{i=1}^k (z_i^2 + 1)^{-1}.$$

Clearly, (2.13) is true with f replaced by g . Moreover, since $g(z_1, \dots, z_k) < e^{-1}$ outside of a sufficiently large sphere about the origin, and since $-f \log f < -g \log g$ if $0 < f < g < e^{-1}$, it suffices to prove (2.14) with f replaced by g , assuming g bounded and (2.13) with f replaced by g . By (2.13), we have

$$(2.16) \quad \int_{R^k} g \log \prod_{i=1}^k (1 + z_i^2)^{\frac{1}{2}} d\mu < \infty.$$

Thus, it suffices to prove the finiteness of

$$(2.17) \quad \begin{aligned} & - \int_{R^k} g \log g d\mu - \int_{R^k} g \log \prod_{i=1}^k (1 + z_i^2)^{\frac{1}{2}} d\mu \\ & = \int_{R^k} g \log \prod_{i=1}^k (1 + z_i^2)^{\frac{1}{2}} \left\{ \frac{-\log [g \prod_{i=1}^k (1 + z_i^2)^{\frac{1}{2}}]}{\log \prod_{i=1}^k (1 + z_i^2)^{\frac{1}{2}}} \right\} d\mu. \end{aligned}$$

The fact that $g(z_1, \dots, z_k) \leq \prod_{i=1}^k (z_i^2 + 1)^{-1}$ (see (2.15)) implies easily that the bracketed expression in (2.17) is ≤ 1 ; by (2.16), this completes the proof of the lemma.

3. Example 1. Structural location parameter, incidental scale parameter.

Let k be a positive integer, let μ be Lebesgue measure on Euclidean k -space, let g be a univariate probability density function with respect to Lebesgue measure, and let

$$(3.1) \quad f(x_i | \theta, \alpha_i) = \frac{1}{\alpha_i^k} \prod_{j=1}^k g\left(\frac{x_{ij} - \theta}{\alpha_i}\right),$$

where $x_i = (x_{i1}, \dots, x_{ik})$. (Thus, observations are taken in groups of $k \geq 1$, the value of the incidental parameter being the same within each group. The (unconditional) density of $X_i = (X_{i1}, \dots, X_{ik})$ is given by Equation (2.1). Thus, the X_i are independent, but, for fixed i , the $X_{ij} (j = 1, \dots, k)$ need not be independent.) Here Ω is the real line. Some further assumptions on g will be made below; we remark here that the important case

$$(3.2) \quad g(x) = (2\pi)^{-1/2} e^{-(x^2/2)}$$

will satisfy our assumptions. (See also (3.4) below.)

The cases $k = 1$ and $k > 1$ are essentially different. In Example 1a the consistency of the m.l. estimator will be proved for $k = 1$ assuming that A is the set of values $\alpha \geq c$ where c is a known positive constant, and it is pointed out that the property of consistency of the m.l. estimator *does not hold* without this assumption. The proof of consistency in Example 1a is actually carried out for $k \geq 1$ since this requires little additional space and will save space in Example 1b where we may refer back to 1a for proofs. In Example 1b we prove consistency of the m.l. estimator in the case $k > 1$ without assuming $\alpha \geq c > 0$.

Example 1a. We assume that $k \geq 1$ and that A is the set of all real values $\alpha \geq c$ where c is a known positive constant. In the case $k = 1$, this assumption on A can be weakened slightly to an assumption on the behavior of $G(\alpha)$ as $\alpha \rightarrow 0$; however, some such assumption is necessary for consistency, since the last example of Section 6 shows that, even in cases where Γ is restricted to a simple parametric class of df's on a set of positive reals which is *not* bounded away from zero, it can happen that no m.l. or modified m.l. estimator exists and that there are neighborhood m.l. estimators which are not consistent.

We now state our assumptions on g and G_0 . They seem reasonable and are in a form which makes brief proofs possible; they undoubtedly can be weakened. (These last remarks apply also to Examples 2 and 3. See also the first part of Section 6 for one method by which we can prove the results of our examples under weaker conditions.) We hereafter assume

- (a) $\sup_x g(x) < \infty$;
- (b) g is lower semicontinuous and for every $\epsilon > 0$ there is a continuous function $h_\epsilon \geq g$ for which $\int [h_\epsilon(x) - g(x)] dx < \epsilon$;
- (c) $\lim_{|x| \rightarrow \infty} g(x) = 0$;
- (3.3) (d) $-\int_{-\infty}^{\infty} g(x) [\log |x|]^+ dx > -\infty$;
- (e) $\int_{-\infty}^{\infty} |x|^t g(x) dx \neq 0$ for almost all real t ;
- (f) $g(x) > 0$ for almost all x in some open interval whose closure contains the point $x = 0$.

We note that, in addition to being satisfied in the case (3.2), Assumption (3.3) is also satisfied in such important cases as

- (a) $g(x) = 1/\pi(1 + x^2)$;
 (3.4) (b) $g(x) = 1$ if $|x| < \frac{1}{2}$ and $g(x) = 0$ otherwise;
 (c) $g(x) = e^{-x}$ if $x > 0$ and $g(x) = 0$ otherwise.

Of course, if g does not satisfy (3.3) but if there is a function g^* satisfying (3.3) and for which $g(x) = g^*(x)$ almost everywhere, then we may carry out our considerations replacing g by g^* .

We assume that Γ consists of all G such that

$$(3.5) \quad \int_c^\infty (\log \alpha) dG(\alpha) < \infty,$$

where c is the constant used before in the definition of A . For example, G belongs to Γ if, for some positive constants b and ϵ ,

$$(3.6) \quad 1 - G(\alpha) < \frac{b}{\log \alpha (\log \log \alpha)^{1+\epsilon}}$$

for $\alpha > e^e$; integration by parts will verify that (3.6) implies (3.5). Condition (3.5) is weaker than the requirement that any positive (not necessarily integral) movement of G be finite.

We now verify the assumptions of Section 2. We complete the definition of f for (θ, α) in $\bar{\Omega} \times \bar{A}$ by setting $f(x | \theta, \alpha) = 0$ whenever $\theta = \pm \infty$ or $\alpha = \infty$. For $(\theta, G) \in \bar{\Omega} \times \bar{\Gamma}$, we then define $f(x | \theta, G)$ by (2.1). (We remark that $\bar{\Gamma}$ obviously contains all df's on \bar{A} .) Assumption 1 is obviously satisfied. Assumption 3 follows from the fact that (3.3) implies that $f(x | \theta, G)$ is for each x lower semicontinuous in (θ, G) (in the sense of the metric δ) on $\bar{\Omega} \times \bar{\Gamma}$, and the fact that $\bar{\Omega} \times \bar{\Gamma}$ is separable. Write $h_i(x_i | \theta, \alpha) = \alpha^{-k} \prod_{j=1}^k h_j[(x_{ij} - \theta)/\alpha]$. In order to verify Assumption 2, we note that, by the lower semicontinuity in (θ, G) of $f(x | \theta, G)$ and by the Helly-Bray theorem, we have (assuming, as we may, that the h_i of (3.3) (b) satisfies $\lim_{|x| \rightarrow \infty} h_i(x) = 0$) that $(\theta_i, G_i) \rightarrow (\theta^*, G^*)$ as $i \rightarrow \infty$ implies

$$(3.7) \quad \begin{aligned} f(x | \theta^*, G^*) &\leq \liminf_{i \rightarrow \infty} \int f(x | \theta_i, \alpha) dG_i \leq \limsup_{i \rightarrow \infty} \int f(x | \theta_i, \alpha) dG_i \\ &\leq \lim_{i \rightarrow \infty} \int h_i(x | \theta_i, \alpha) dG_i = \int h_i(x | \theta^*, \alpha) dG^*. \end{aligned}$$

Since the last member of (3.7) is greater than or equal to the first for all x and since their difference has integral $< \epsilon^k$ (with respect to μ), Assumption 2 follows at once.

In verifying Assumption 4, it clearly suffices to prove that, if $f(x | \theta_0, G_0) =$

$f(x | \theta_i, G_i)$ for almost all x , where $(\theta_i, G_i) \in \Omega \times \Gamma$ for $i = 0, 1$, then $(\theta_0, G_0) = (\theta_1, G_1)$. If an interval $0 < x < \epsilon$ satisfies (3.3) (f), there is a value β such that

$$P\{X_{1j} \leq t \text{ for } 1 \leq j \leq k | \theta_0, G_0\} \leq \beta$$

is satisfied (whatever be G_0) if and only if $t \leq \theta_0$, a similar assertion holding if the interval $-\epsilon < x < 0$ satisfies (3.3) (f). Hence, it suffices to prove the above assertion when $\theta_0 = \theta_1$, since it cannot hold when $\theta_0 \neq \theta_1$. Let H_i be the df of the random variable $\log \alpha_i$ when G_i is the df of the random variable α_i ; i.e., $H_i(t) = G_i(e^t)$. Then, putting $g^*(z) = e^z[g(e^z) + g(-e^z)]$ (g^* is the density of $\log |U|$ when g is the density of U), it suffices to prove that, if H_0 and H_1 are not identical, then $p_1(z_1, \dots, z_k)$ and $p_2(z_1, \dots, z_k)$ are not identical for almost all (z_1, \dots, z_k) , where

$$(3.8) \quad p_i(z_1, \dots, z_k) = \int_{-\infty}^{\infty} \prod_{j=1}^k g^*(z_j - \beta) dH_i(\beta).$$

Let g^{**} be the density function of $\sum_1^k Z_j/k$ when the Z_j are independent random variables with common density g^* . The above assertion is then implied by the assertion that the function

$$(3.9) \quad q(r) = \int_{-\infty}^{\infty} g^{**}(r - \beta) dH(\beta)$$

uniquely determines the df H . But if A, B, C are the characteristic functions of g, g^{**}, H , respectively, then $B(t) \neq 0$ for almost all t by (3.3) (e) and hence $C(t)$ is determined for those t for which $B(t) \neq 0$ by $C(t) = A(t)/B(t)$ and elsewhere by continuity. Thus, Assumption 4 is verified.

It remains to verify Assumption 5. Since $f(x | \theta, G)$ is uniformly bounded in x, θ, G , Assumption 5 will clearly be satisfied if

$$(3.10) \quad E \log f(X_1 | \theta_0, G_0) > -\infty.$$

Since the left side of (3.10) does not depend on θ_0 , we may assume $\theta_0 = 0$. By (3.3) (d) and (3.5), we have

$$(3.11) \quad \begin{aligned} E[\log |X_{11}|]^+ &= E \left[\log \frac{|X_{11}|}{\alpha_1} + \log \alpha_1 \right]^+ \\ &\leq E \left[\log \frac{|X_{11}|}{\alpha_1} \right]^+ + E[\log \alpha_1]^+ < \infty; \end{aligned}$$

equation (3.10) is a consequence of (3.11) and the lemma at the end of Section 2.

This completes our verification of the fact that the assumptions of Section 2 are implied by (3.3) and (3.5).

Example 1b. We now assume $k > 1$. A is the set of all positive α , while Γ is the set of all df's G on A satisfying

$$(3.12) \quad \int_0^{\infty} |\log \alpha| dG(\alpha) < \infty.$$

We assume that g satisfies (3.3) (some alterations could be made here but, for the sake of brevity, we forego making them) and also that

$$(3.13) \quad \begin{aligned} (a) \quad & \lim_{|x| \rightarrow \infty} xg(x) = 0; \\ (b) \quad & \sup_{x_1} [\min_{r < j} |x_{1r} - x_{1j}|]^k \prod_{j=1}^k g(x_{1j}) < \infty. \end{aligned}$$

Assumption (3.13) is easily verified, for example, in cases (3.2) and (3.4).

We now verify the assumptions of Section 2. We define $f(x | \theta, \alpha) = 0$ whenever $\theta = \pm \infty$ or $\alpha = 0$ or ∞ ; $f(x | \theta, G)$ is then defined by (2.1) for $(\theta, G) \in \bar{\Omega} \times \bar{\Gamma}$. Assumptions 1, 3, and 4 are verified exactly as in Example 1a. In verifying Assumption 2, we may follow the demonstration of Example 1a, noting only that the h_i of (3.3) (b) may (because of (3.13) (a)) clearly be assumed to satisfy $\lim_{|x| \rightarrow \infty} xh_i(x) = 0$, so that for every x none of whose components is θ^* ,

$$(3.14) \quad \lim_{\substack{i \rightarrow \infty \\ \alpha \rightarrow 0}} h_i(x | \theta_i, \alpha) = 0;$$

thus, for almost all (μ) x , the Helly-Bray theorem may still be used at the last step of (3.7), no difficulty being caused by the possibility that $\liminf_{i \rightarrow \infty} G_i(0) < G^*(0)$.

It remains to verify Assumption 5. Now, $f(x | \theta, G)$ is no longer uniformly bounded as it was in Example 1a. However, by (3.13) (b), there is a constant B such that, for all $x_1 = (x_{11}, \dots, x_{1k})$ none of whose components are equal, every $\theta \in \Omega$, and every $\alpha \in A$,

$$(3.15) \quad \begin{aligned} f(x_1 | \theta, \alpha) &= [\min_{r < s} |x_{1r} - x_{1s}|]^{-k} \left\{ [\min_{r < s} |y_{1r} - y_{1s}|]^k \prod_{j=1}^k g(y_j) \right\} \\ &\leq B [\min_{r < s} |x_{1r} - x_{1s}|]^{-k}, \end{aligned}$$

where $y_{1r} = (x_{1r} - \theta)/\alpha$. Hence, for almost all x_1 ,

$$(3.16) \quad \begin{aligned} \sup_{\substack{\theta \in \Omega \\ \alpha \in A}} \log f(x_1 | \theta, \alpha) &\leq \log B + k \max_{r < s} \log [1/|x_{1r} - x_{1s}|] \\ &\leq \log B + k \sum_{\substack{r, s \\ r < s}} [\log (1/|x_{1r} - x_{1s}|)]^+. \end{aligned}$$

Now, by (3.3) (a), there is a value B' such that $g(z) \leq B'$ for all z . Hence, by (3.12), B_1 denoting a finite constant, we have

$$(3.17) \quad \begin{aligned} E[\log(1/|X_{11} - X_{12}|)]^+ &\leq E[\log 1/\alpha]^+ + E[\log(\alpha/|X_{11} - X_{12}|)]^+ \\ &\leq B_1 - 2 \int_{-\infty}^{\infty} g(z_2) \int_{z_2}^{z_2+1} B' \log(z_1 - z_2) dz_1 dz_2 \\ &= B_1 + 2B' < \infty. \end{aligned}$$

From (3.16) and (3.17), we obtain

$$(3.18) \quad E \sup_{\gamma \in \bar{\Omega} \times \bar{\Gamma}} \log f(X_1 | \gamma) < \infty.$$

Assumption 5 is a consequence of (3.18) and of (3.10), the latter of which is proved exactly as in Example 1a. This completes the verification of the assumptions of Section 2 in Example 1b.

The discrete analogue of Example 1 can be carried out similarly by letting x , θ , α take on only rational values; this is, however, of less practical importance. The multivariate extension of Example 1 (X_{ij} a vector) may also be carried out similarly.

4. Example 2. The straight line with both variables subject to error.

In this section we shall treat the case $k = 1$ of fitting a straight line with both variables subject to normal error, a famous problem with a long history.

We consider a system $\{(X_{i1}, X_{i2})\}$, $i = 1, 2, \dots$, of independent chance 2-vectors (the two components X_{i1} , X_{i2} need not be independent for fixed i). We have $\theta = (\theta_1, \theta_2)$, Ω the entire plane, $\theta_0 = (\theta_{10}, \theta_{20})$, A the entire line. Γ is the totality of all non-normal (univariate) distributions G (a chance variable which is constant with probability one is to be considered normally distributed with variance zero) which satisfy

$$\int (\log |\alpha|)^+ dG(\alpha) < \infty.$$

It is known to the statistician that

$$\begin{aligned} X_{i1} &= \alpha_i + u_i, \\ X_{i2} &= \theta_{10} + \theta_{20}\alpha_i + v_i, \end{aligned}$$

where (u_i, v_i) are jointly normally distributed chance variables with means zero, each pair (u_i, v_i) distributed independently of every other pair and of the independent chance variables $\{\alpha_i\}$, with a common covariance matrix which is unknown to the statistician.

It is known (see [10]) that the distribution of (X_{i1}, X_{i2}) then determines θ_0 uniquely, but in general not G_0 , the "true" df of α_i , or the "true" covariance matrix

$$\begin{Bmatrix} d_{11}^0 & d_{12}^0 \\ d_{12}^0 & d_{22}^0 \end{Bmatrix}$$

of (u_i, v_i) . However, a "canonical" complex is determined. (See [4].)

Complete the spaces Ω , A , and Γ to obtain $\bar{\Omega}$, \bar{A} and $\bar{\Gamma}$. The space $\bar{\Gamma}$ contains all normal distributions on A , but this will cause us no trouble in estimating θ_0 , as we shall soon see.

Let D be the space of all triples (d_{11}, d_{12}, d_{22}) such that

$$d_{11} \geq \lambda_{11} > 0, \quad d_{22} \geq \lambda_{22} > 0,$$

$$d_{11} d_{22} - d_{12}^2 \geq \lambda_{12} > 0,$$

where $\lambda_{11}, \lambda_{12}, \lambda_{22}$, are given positive numbers. (This will be discussed further below.) We define a metric in D in the same way that one is defined on Ω . Let \bar{D} be the completed space. We shall assume that the "true" triple $d_{11}^0, d_{12}^0, d_{22}^0$ is in D .

The place of $\bar{\Omega} \times \bar{\Gamma}$ in Section 2 and in Example 1 will now be taken by $\bar{\Omega} \times \bar{\Gamma} \times \bar{D}$. We therefore define

$$\gamma = (\theta_1, \theta_2, G, d_{11}, d_{12}, d_{22})$$

as the generic point in $\bar{\Omega} \times \bar{\Gamma} \times \bar{D}$.

Let $f(x_1, x_2 | \theta_1, \theta_2, \alpha, d_{11}, d_{12}, d_{22})$ be the joint density function of (X_{a1}, X_{a2}) when $\theta = (\theta_1, \theta_2)$, $\alpha_i = \alpha$, and the covariance matrix of (u_i, v_i) is

$$\begin{Bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{Bmatrix}$$

(μ is Lebesgue measure in the plane). If, in the above, θ is in $\bar{\Omega} - \Omega$ or α is in $\bar{A} - A$ or (d_{11}, d_{12}, d_{22}) is in $\bar{D} - D$, we define f to be zero. Finally we define

$$f(x_1, x_2 | \gamma) = \int_A f(x_1, x_2 | \theta_1, \theta_2, \alpha, d_{11}, d_{12}, d_{22}) dG(\alpha).$$

It is known ([10] and [4]) that all γ in the same canonical class, and only such, define the same $f(x_1, x_2 | \gamma)$ (of course, to within a set of μ -measure zero). Two members of the same canonical class have the same $\theta = (\theta_1, \theta_2)$ but different G 's and d_{ij} 's. We shall estimate only θ_0 . For an estimator of the entire canonical complex by the minimum distance method under necessary assumptions only, see [4].⁵ In Section 5 below will be found an explanation of why the entire canonical complex cannot be estimated by the m.l. method.

From the definition of $f(x_1, x_2 | \gamma)$ it follows immediately that Assumptions 1, 2, and 3 of Section 2 are satisfied. Since we are estimating only θ_0 , it is sufficient to verify Assumption 4 only for θ_0 and $\theta^* \neq \theta_0$, i.e., if we write the γ_0 and γ_1 of (2.4) as

$$\gamma_0 = (\theta_{10}, \theta_{20}, G_0, d_{11}^0, d_{12}^0, d_{22}^0),$$

$$\gamma_1 = (\theta_1^*, \theta_2^*, G_1, d_{11}, d_{12}, d_{22}),$$

only $\theta_0 = (\theta_{10}, \theta_{20})$ has to be different from the corresponding $\theta^* = (\theta_1^*, \theta_2^*)$. Now we know that G_0 is in Γ , hence is not normal and assigns probability one to A . If G_1 is also in Γ then Assumption 4 follows at once from the results of

⁵ See footnote 4.

Reiersøl [10] or from [11]. If G_1 assigns probability less than one to A , $f(x | \gamma_1)$ assigns probability less than one to the Euclidean plane of (x_1, x_2) . If G_1 is normal and assigns probability one to A , then (X_{11}, X_{12}) are jointly normal under γ_1 , but not under γ_0 . Thus Assumption 4 is always satisfied.

To verify Assumption 5 we proceed essentially as in Example 1, and use the lemma at the end of Section 2. Assumption 5 is satisfied if

$$E \log f(X_{11}, X_{12} | \gamma_0) > -\infty.$$

By the lemma this will follow if we prove

$$E\{\log |X_{1j}|\}^+ < \infty$$

for $j = 1, 2$. Now

$$\begin{aligned} E\{\log |X_{11}|\}^+ &\leq E\{\log [|X_{11} - \alpha_1| + |\alpha_1|]\}^+ \\ &\leq E\{\log [|X_{11} - \alpha_1| + 1]\} + E\{\log |\alpha_1|\}^+ \\ &= E\{\log [|u_1| + 1]\} + E\{\log |\alpha_1|\}^+ \\ &< \infty. \end{aligned}$$

Similarly,

$$\begin{aligned} E\{\log |X_{12}|\}^+ &\leq E\{\log [|X_{12} - \theta_{10} - \theta_{20}\alpha_1| + |\theta_{10} + \theta_{20}\alpha_1|]\}^+ \\ &\leq E \log [|X_{12} - \theta_{10} - \theta_{20}\alpha_1| + 1] + E\{\log |\theta_{10} + \theta_{20}\alpha_1|\}^+ \\ &\leq E \log [|v_1| + 1] + \{\log |\theta_{10}|\}^+ + E \log [1 + |\theta_{20}\alpha_1|] \\ &< \infty. \end{aligned}$$

Thus we have shown, under our assumptions on Γ and D , that Assumptions 1 through 5 of Section 2 are satisfied, so that the m.l. estimator of θ_0 converges strongly to θ_0 as $n \rightarrow \infty$.

The assumption on D (that d_{11} , d_{22} , and $d_{11}d_{22} - d_{12}^2$ are bounded away from zero) cannot be entirely dispensed with. For if D consists of all triples for which d_{11} , d_{22} , and $d_{11}d_{22} - d_{12}^2$ are positive, if S_n is the sample df of x_{11}, \dots, x_{n1} , and if $\hat{\gamma}_n$ is the complex $(0, 0, S_n, \epsilon, 0, \sum_1^n x_{12}^2)$, then it is easily verified that $\lim_{n \rightarrow \infty} L((x_{11}, x_{12}), \dots, (x_{n1}, x_{n2}) | \hat{\gamma}_n) = \infty$; thus, no m.l. or modified m.l. estimator exists, and there are neighborhood m.l. estimators which are not consistent (for θ).

The case $k > 1$ is much simpler to treat than the above case. It is easy to see that then the covariance matrix of (u_i, v_i) is uniquely determined, and from this it follows easily that the whole complex γ is uniquely determined. The problem can be treated in a manner similar to that of Examples 1b and 3b.

The problem of this section with the distribution of (u_i, v_i) other than normal may also be treated by the m.l. method, as in Examples 1 and 3. The last paragraph of Section 3 applies also to the present example.

5. Example 3. Structural scale parameter, incidental location parameter.

We consider here the case of a structural scale parameter and an incidental location parameter; this reverses the roles of the two parameters of Example 1. Thus, we suppose μ to be Lebesgue measure on R^k and

$$(5.1) \quad f(x_i | \theta, \alpha) = \frac{1}{\theta^k} \prod_{j=1}^k g\left(\frac{x_{ij} - \alpha}{\theta}\right).$$

The cases $k = 1$ and $k > 1$ are essentially different, and we consider them separately.

Example 3a. The case $k = 1$. This example is another simple one where no m.l. estimator is consistent, and also shows, in a simpler setting, why in Example 2 the m.l. method was incapable of estimating the components of the canonical complex other than θ . Since Example 3a is intended to illustrate the *failure* of the m.l. method in certain situations, we shall for simplicity assume that g is given by (3.2); examples with other g (e.g., (3.4)) may be treated similarly. Ω may be taken to be any specified set of positive numbers containing more than one point; for the sake of brevity, we assume that Ω contains its greatest lower bound c (say) (and thus, that $c > 0$), but it is easy to carry through a similar demonstration (with modified or neighborhood m.l. estimators in place of m.l. estimators) when $c \notin \Omega$. Γ is taken to be the class of all df's G on the real line for which $\int [\log |\alpha|]^+ dG(\alpha) < \infty$ and such that G has no normal component; i.e., no G in Γ can be represented as the convolution of two df's, one of which is normal with positive variance. (Γ may be further restricted, e.g., by the condition that for each G there is a bounded set outside of which G has no variation.)

All assumptions of Section 2 are easily verified except Assumption 4; there is no difficulty of identifiability in $\Omega \times \Gamma$, but there clearly *is* in $\bar{\Omega} \times \bar{\Gamma}$. Consider now the expression

$$(5.2) \quad \prod_{i=1}^n \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2} c} e^{-(1/2c^2)(x_i - s)^2} dM(s).$$

It is clear that the maximum of (5.2) with respect to M can be achieved only by an M which assigns probability one to the interval $(\min(x_1, \dots, x_n), \max(x_1, \dots, x_n))$ and hence which has no normal component. This discussion of the expression (5.2) shows that, for every n , any m.l. estimator (the fact that the maximum is attained is easily verified) of (θ, G) subject to our assumption $\theta \geq c$ *always* estimates θ to be c . Thus, no m.l. estimator of (θ, G) is consistent (unless $\theta = c$).

To summarize the result of this example, then, the m.l. method is incapable of estimating consistently the normal component of the df of the sequence $\{X_i\}$ of independent identically distributed random variables because, in every neighborhood of a point (θ, G) with $\theta > c$, there are points with $\theta = c$ (and for which the likelihood is larger).

Let N_* denote the normal df with mean 0 and variance σ^2 , and let $H_1 * H_2$ denote the convolution of the two df's H_1 and H_2 .

It is interesting to note that, without any assumption on Γ (except the necessary identifiability assumption that G_0 has no normal component), the minimum distance method is capable of estimating (θ_0, G_0) consistently [4]. The difficulty noted above for the m.l. estimator is avoided by noting the *rate* at which the sample df S_n converges to the df $N_{\theta_0} * G_0$ of X_1 and estimating θ_0 *not* by the value t for which $N_t * H$ is closest to S_n for some normal-free H (this would encounter the same difficulty as the m.l. estimator, since, the smaller t is taken, the closer can $N_t * H$ be made to approximate S_n), but as the *largest* value for which there is an $N_t * H$ suitably close to S_n ("suitably" is connected with the rate mentioned above.)

One could modify the example as considered above so as not to require G_0 to have no normal component, and try then to escape the difficulty of non-identifiability by asking for an estimator of the canonical representation of (θ, G) , this representation consisting of two df's, the normal and nonnormal components of $N_\theta * G$. The previous demonstration then shows that no m.l. estimator of the canonical representation estimates it consistently, and thus illustrates, in a simpler setting than that of Example 2 with $k = 1$, why the m.l. estimator could not be used in Example 2 to estimate the components of the canonical complex other than θ .

We remark that it is easy in many cases such as that of the present example to prove a result such as the one that, (t_n, H_n) denoting an m.l. estimator of (θ_0, G_0) after n observations, the df $N_{t_n} * H_n$ converges w.p.1 to $N_{\theta_0} * G_0$ as $n \rightarrow \infty$. Such a property is much weaker than that of the consistency of the m.l. estimator, and does not lie much deeper than the Glivenko-Cantelli theorem.

Example 3b. The case $k > 1$. We assume f to be given by (5.1) with $k > 1$. The function g is assumed to satisfy the conditions (a), (b), (c), and (d) of (3.3); conditions (a) and (b) of (3.13), and

$$(5.3) \quad \int_{-\infty}^{\infty} e^{itz} g(x) dx \neq 0 \quad \text{for almost all real } t.$$

(As in Example 1a, weaker conditions could be assumed here if we assumed also $\theta \geq c > 0$; the above conditions are analogous to those of Example 1b.) Thus, for example, (3.2) and (3.4) satisfy these assumptions. Ω is the set of all values $\theta > 0$, while A is the real line and Γ is the set of all df's G on A for which

$$(5.4) \quad \int_{-\infty}^{\infty} |\log |\alpha||^+ dG(\alpha) < \infty.$$

We now verify the assumptions of Section 2. We define $f(x | \theta, \alpha) = 0$ when $\theta = 0$ or ∞ or $\alpha = \pm \infty$. The definition of $f(x | \theta, G)$ for $(\theta, G) \in \bar{\Omega} \times \bar{\Gamma}$ is then given by (2.1). Assumptions 1, 2, and 3 are now verified as in Example 1b, interchanging the roles of θ and α in the latter (including the definition of $h_\mu(x | \theta, \alpha)$) and noting that (3.14) still holds for almost all $(\mu) x$, with this interchange. In

order to verify Assumption 4, we note, for $(\theta, G) \in \Omega \times \Gamma$, that θ is determined by the density function of $X_{11} - X_{12}$ and that, for almost all real t , the characteristic function of G is then given by $B(t/k, \dots, t/k)/[C(\theta t/k)]^k$ where $B(t_1, \dots, t_k)$ is the characteristic function of X_{11}, \dots, X_{1k} and $C(t)$ is the characteristic function of g .

Finally, Assumption 5 is a consequence of equation (3.18), which is proved in the present case exactly as in Example 1b (using (3.15), (3.16), and (3.17), with α_1 replaced by θ in the latter), and of equation (3.10) (with f defined by (5.1)). Equation (3.10) in the present example is a consequence of the lemma at the end of Section 2 and of

$$(5.5) \quad \begin{aligned} E\{\log |X_{11}|\}^+ &\leq E\{\log [|X_{11} - \alpha_1| + |\alpha_1|]\}^+ \\ &\leq E\log [|X_{11} - \alpha_1| + 1] + E\{\log |\alpha_1|\}^+ < \infty. \end{aligned}$$

This completes the verification of the assumptions of Section 2 in Example 3b. The last paragraph of Section 3 applies also to the present example.

6. The Classical case. Miscellaneous remarks. It does not seem to have been noticed in the literature that a simple device exists for proving consistency of the m.l. estimator in certain cases where the regularity conditions of published proofs fail. This device may be used in the case studied in the present paper (to prove consistency in the examples under weaker conditions than those stated) as well as in the classical parametric case. We now illustrate this device in an example of the latter case.

When Γ consists only of distributions which give probability one to a single point, the problem of the present paper becomes the classical problem of estimating the parameter θ and the parameter σ (say) to which G_0 gives probability one. If θ may be any real value and σ any positive value, then the function $(1/\sigma)g((x - \theta)/\sigma)$ of Section 3 does not satisfy Wald's integrability condition or the corresponding condition of any other published proof; one verifies easily that (2.5) is not satisfied for any point in the (θ, σ) half-plane which lies on the line $\sigma = 0$. (The line $\sigma = 0$ has to be added to Ω in the process of forming $\bar{\Omega}$. As in earlier sections, we assume the true σ_0 to be > 0 .) Often, however, when the observations are considered as if they were taken in groups of two or more, the integrability condition will be satisfied. Such is the case, for example, with the density function

$$\frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x_1 - \theta)^2} \cdot \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x_2 - \theta)^2}$$

and the normal density function

$$\frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{1}{2} \frac{(x_1 - \theta)^2}{\sigma^2}\right\} \cdot \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{1}{2} \frac{(x_2 - \theta)^2}{\sigma^2}\right\}.$$

(Of course the estimator from the normal distribution is known to be consistent, but this does not alter the validity of the example.) In such cases it

follows from Wald's proof [5] (using the compactification device used above) or from the result of Example 1b that the m.l. sequence of estimators considered only after an even number of observations is consistent, and from this it is an easy matter to show that the entire m.l. sequence of estimators is consistent.

We shall now discuss the integrability conditions of [5] and of the present paper. The integrability condition (2.5) involves the difference of two logarithms; the integrability condition as given by Wald in [5] requires the finiteness of the expected value of each logarithm. The form (2.5) is satisfied whenever the condition of [5] is, and has one other advantage which we shall now illustrate by an example. Let the observed chance variable X have density function $\theta e^{-\theta x}$ for $x > 0$ and zero elsewhere. The parameter θ is unknown and Ω is the positive half-line, so that $\bar{\Omega}$ contains the point $\theta = 0$. One verifies easily that the condition of [5], and hence (2.5), are satisfied. Suppose now that, instead of observing X , one observes $Y = e^{(e^X)}$, which therefore has the density function

$$\frac{\theta}{x} (\log x)^{-\theta-1}$$

for $x > e$, and zero elsewhere. One readily verifies that, when $\theta < 1$,

$$E \log \left\{ \frac{\theta}{Y} (\log Y)^{-\theta-1} \right\} = -\infty,$$

so that the condition of [5] is not satisfied when $0 < \theta_0 < 1$. Thus, whether the condition of [5] is satisfied depends in this instance on whether one observes X or Y ; this is an unfortunate circumstance, since the estimation problems are in simple correspondence. On the other hand, condition (2.5) is invariant under one-to-one transformation of the observed chance variable because the numerator and denominator of the ratio in (2.5) are multiplied by the same Jacobian. (In particular, therefore, the chance variable Y satisfies (2.5).)

Without resorting to artificial or pathologic examples as is sometimes done in the literature, it is still easy to give instances where the m.l. method does not give consistent estimators in the classical parametric case. For example, consider the density function

$$\frac{1}{2(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2}(x - \theta)^2 \right\} + \frac{1}{2(2\pi)^{1/2}\sigma} \exp \left\{ -\frac{1}{2} \frac{(x - \theta)^2}{\sigma^2} \right\}$$

of the sequence of independent and identically distributed chance variables X_1, X_2, \dots . Here θ and σ are the unknown parameters, θ may be any real number and σ any positive number. It is easy to see that the supremum of the likelihood function is almost always infinite, no m.l. or modified m.l. estimator exists, and there are neighborhood m.l. estimators (where θ_0 is estimated by X_1 , say) which are obviously not consistent.

REFERENCES

- [1] J. NEYMAN AND E. L. SCOTT, "Consistent estimates based on partially consistent observations" *Econometrica*, Vol. 16 (1948), pp. 1-32.

- [2] A. WALD, "The fitting of straight lines if both variables are subject to error," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 284-300.
- [3] A. DVORETZKY, J. KIEFER AND J. WOLFOWITZ, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 642-669.
- [4] J. WOLFOWITZ, "Estimation of the components of stochastic structures," *Proc. Nat. Acad. Sci., U.S.A.*, Vol. 40, No. 7 (1954), pp. 602-606.
- [5] A. WALD, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 595-601.
- [6] A. WALD, "Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 220-227.
- [7] J. NEYMAN, "Existence of consistent estimates of the directional parameter in a linear structural relation between two variables," *Ann. Math. Stat.* Vol. 22 (1951), pp. 497-512.
- [8] J. NEYMAN AND E. L. SCOTT, "On certain methods of estimating the linear structural relation between two variables," *Ann. Math. Stat.* Vol. 22 (1951), pp. 352-361.
- [9] J. WOLFOWITZ, "Estimation by the minimum distance method," *Ann. Inst. Stat. Math. Tokyo*, Vol. 5 (1953), pp. 9-23.
- [10] O. REIERSØL, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375-389.
- [11] J. WOLFOWITZ, "Consistent estimators of the parameters of a linear structural relation," *Skand. Aktuarietids.* (1952), pp. 132-151.
- [12] J. WOLFOWITZ, "On Wald's proof of the consistency of the maximum likelihood estimate," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 602-603.

AVERAGE VALUES OF MEAN SQUARES IN FACTORIALS¹

BY JEROME CORNFIELD AND JOHN W. TUKEY

National Institutes of Health and Princeton University

1. Summary. The assumptions appropriate to the application of analysis of variance to specific examples, and the effects of these assumptions on the resulting interpretations, are today a matter of very active discussion. Formulas for average values of mean squares play a central role in this problem, as do assumptions about interactions. This paper presents formulas for crossed (and, incidentally, for nested and for non-interacting completely randomized) classifications, based on a model of sufficient generality and flexibility that the necessary assumptions concern only the selection of the levels of the factors and not the behavior of what is being experimented upon. (This means, in particular, that the average response is an arbitrary function of the factors.) These formulas are not very complex, and specialize to the classical results for crossed and nested classifications, when appropriate restrictions are made.

Complete randomization is only discussed for the elementary case of "no interactions with experimental units" and randomized blocks are not discussed. In discussion and proof, we give most space to the two-way classification with replication, basing our direct proof more closely on the proof independently obtained by Cornfield [17], than on the earlier proof by Tukey [20]. We also treat the three-way classification in detail. Results for the general factorial are also stated and proved.

The relation of this paper to other recent work, published and unpublished, is discussed in Section 4 (average values of mean squares) and in Section 11 (various types of linear models).

INITIAL DISCUSSION

2. Introduction. During the last years of the last decade it was relatively easy to believe that the analysis of variance was well understood. Eisenhart's summary article of 1947 [5], when combined with the work of Pitman [13] and Welch [15] on the randomization approach (work published in 1937-1938, which ever since has been far too much neglected), seemed to provide a simple, easily understandable account of the foundations. But as the years have passed, both statisticians and users of analysis of variance have gradually become aware of a number of areas in which we needed to deepen our understanding. One of these is the relation of formulas for average values of mean squares to assumptions. These are of central importance, since the choice of an "error term" as a basis for *either*

Received May 31, 1955; revised May 24, 1956.

¹ Based in part on the Statistical Research Group, Princeton University, Memorandum Report 18, "Interaction in a row-by-column design," prepared in connection with research sponsored by the Office of Naval Research.

significance tests or confidence statements *must at least* take into account average values of mean squares. (It would be very desirable to consider, also, variances, covariances, and distributions of mean squares, particularly if one is concerned with the detailed validity of an F -test or a multiple-comparisons procedure. Average values, however, seem absolutely essential.) These average values should apply when there are real effects and should not be confined to so called "null" hypotheses. To derive them we must make assumptions, but they should be as weak as possible.

The present paper deals with one aspect of the problem of average mean squares. After introducing this aspect, and describing the situation which we are to treat, we shall relate both this aspect to other aspects of current interest and our work to the work of others.

Difficulties in applying the analysis of variance to even a two-way model have arisen in the so-called mixed model where rows are sampled and columns are completely enumerated. Of two recent standard textbooks, one concludes that the average value of both row and column mean squares includes a component of variance due to interaction [11]; the other finds such a component in column sum of squares, but not in rows [1]. Both texts assume that observed values are linear combinations of certain fixed and random variables, but differ in the nature of the restrictions that are imposed upon these variables. Granting either text its assumptions, its conclusions necessarily follow.

In analyzing the given body of data, the choice among such assumptions can lead to quite different answers. The choice is thus an important one. But it is not a simple choice. At one stage in the development of his ideas about two-way classifications, one of the authors, who leans toward linear models, had 512 alternative sets of assumptions. If they had led to 512 sorts of analysis, the situation would obviously be quite impractical. The only effective way out of such a difficulty was to obtain a single flexible model which could be specialized to any one of the 512 special possibilities. This was done in [20] (and independently in [17] and in [21]).

The question of what assumptions to make seems, at first glance, to be a purely empirical question, one that should be referred to the subject-matter knowledge of the experimenter, who is the expert on such matters. Sometimes this is helpful and sometimes not. But closer study shows that the choice of assumptions depends on more than empirical questions about the behavior of the experimental material. It depends on the nature of the sampling and randomization involved in obtaining the data (as has been recognized by many statisticians, and recently emphasized by Kempthorne and by Wilk). Moreover, it often depends on the purpose of the analysis, as expressed by the situations or populations to which one wishes to make *statistical* inference. These dependences imply diversity, and adequate treatment of diversity requires flexibility of assumption.

After a general initial discussion (Sections 2-6), various descriptions of the situations treated are given (Sections 7-12). The results (Sections 13-19) come next, and are followed by the proofs (Sections 20-28).

3. A pigeonhole model. These requirements of flexibility are met by the system of assumptions which we are about to describe, which specify an object appropriately called a pigeonhole model. Given the values of each of the factors (variables), we are directed to a pigeonhole containing a finite or infinite population. This population represents the possible results of experimentation with these values of the factors. So far we have made no assumptions about how the factors combine to produce the typical effect of each combination. (This typical effect will often be taken as the arithmetic mean of the population of possible values in the corresponding pigeonhole.) We have made the assumption that we can recognize the different levels of each factor, and we shall shortly make assumptions as to how we sample this array of pigeonholes. Neither of these is an assumption about "how the world behaves"; both are assumptions about the experimenter's behavior (and are consequently much easier to check).

We have emphasized the generality of our assumptions; we must also emphasize their limitations. We are concerned with situations where the variability of experimental units (plots, reactors, epochs, mice, etc.) does not play such a predominant role as to require special attention in the design of the experiment. We shall not attempt to treat such well-known situations as randomized blocks, which have been recently studied by Kempthorne [8] and Wilk [16]. (See the next section for a discussion of mutual relations and distinctions.)

Let us be specific about the case of the two-way classification. Let there be RC pigeonholes arranged in R rows and C columns. Let there be at least n elements in the population in each pigeonhole. Let a sample of r rows be drawn from the R potential rows. Let a sample of c columns be drawn from the C potential columns. The rc intersections of a selected row with a selected column specify the rc pigeonholes which become the cells of the actual experiment. In each of these rc cells, let a sample of n elements be drawn. The values of the rcn elements thus obtained are the numbers which are to be analyzed. *Assume that all the samplings—of rows, of columns, and within pigeonholes—are at random and independent of one another.* This is the only assumption we shall make. Note that it is an assumption about the set-up of the experiment and not about the behavior of those things on which the experiment is performed.

While the generality of this model is quite apparent, its flexibility may not be completely evident. If we choose $R = r$ and $C = c$, then rows and columns become fixed, and we have a fixed model which generalizes Model I of Eisenhart [5], since (i) neither normality or constant variance is assumed for the cells, and (ii) no assumption is made about interactions.

If, at the other extreme, we take R and C both infinite, we have a "random" model which generalizes Model II of Eisenhart, since (i) neither normality or constant variance is assumed for the cells, (ii) normality is not assumed for the row and column populations, and (iii) no assumption is made about interactions.

If we take $C = c$ and R infinite, then we obtain a generalization of the conventional mixed model.

Thus our model is flexible enough to cover all the classical cases, and many others besides.

Our model can be described from two apparently quite different points of view, namely, as an example of urn sampling or in terms of a very general linear model. (By urn sampling we shall mean what is usually referred to as sampling without replacements from a finite population.) We shall give both descriptions, striving to take as different points of view as we can about them. An understanding of either description should suffice as a basis for the interpretation of the results.

4. Relation to other problems and workers. Historically, the understanding of the average mean squares, and of the formulas relating them to the underlying models, has developed separately for the different relations which classifications have to one another in the customary situations. The order of development has usually been the same. First, average mean squares are obtained under simplifying assumptions, and then new formulas are obtained as these assumptions are relaxed, usually successively. The initial assumptions, whose removal is often very important, are typically of the following sorts:

- (1) some classification does not interact with one or more others;
- (2) the "errors" are solely (or practically solely) due to the experimental units used;
- (3a) the levels of a certain variable are *fixed*;
- (3b) the levels of a certain variable are a random sample from an *infinite* population.

When we remove these assumptions, we remove (1) or (2) entirely, and replace (3a) or (3b) by

- (3) the levels of a certain variable are a random sample from a population of arbitrary size.

We shall describe assumption (2) as "no free errors" and assumptions (3a) and (3b) as "fixed—" and "infinite sampled—" respectively.

The foremost distinction between the different relations between classifications is the minimum number of classifications which can enter this relation. Thus two (or more) classifications can be crossed or nested. Three (or more) classifications are involved in randomized blocks, in a simple fractional factorial, or in a Latin square. At least four classifications appear in a lattice, and so on.

First, let us discuss the relations which need involve only two classifications:

(1) *Nested* (or as some say, *hierarchical*) *classifications*. Here the general situation was clarified first, and general formulas have been available for some time. (The early clarification of this case may have been due to its intimate relation to the precision of multistage sampling.)

(2) *Crossed classifications*. The case of random interactions was treated early (cf. [5], [8], etc.), and in varying degrees of generality, the most general being in [14].

For the important case of arbitrary interactions, we know of nothing which antedates Memorandum Report 18 [20], which stated a general rule for crossed and nested cases. (However, informed understanding of the limiting cases of "fixed" and "infinitely sampled" cases seems to have been present among some users before this time.) The rule for the limiting cases was stated by Kempthorne

[8], p. 574). The general results for the two-way classification were obtained by Cornfield [17] and by Wilk [21], independently of each other and of Memorandum Report 18. All three approaches to this result were from somewhat different points of view. Kempthorne ([19], pp. 204-220) has discussed the general two-way situation, basing his treatment on [21]. Bennett and Franklin [2] have sketched a proof for the two-way classification (without replication) (pp. 474-477), and have stated the result for the three-way case (p. 394).

Wilk and Kempthorne [24] have treated the case of the general two-way classification with general sampling of the two factors, with replication, and with complete randomization of experimental units, but without free errors or interactions with experimental units. They have also [23] dealt with the three-way classification, with general proportional numbers in a cell, general sampling of all factors, complete randomization of experimental units, free errors only restricted to be of the same variance, but with no interaction with experimental units. Wilk [25] has continued the analysis of the three-way classification, treating both interaction with experimental units, and, separately, the analysis of cell means for general disproportionate cell numbers. Like the earlier work at Ames, this work was carried out independently of ours, in ignorance of the existence of [20] and [17], and before the appearance of [2].

A short proof, using more special techniques, has been found for the two-way classification by Hooke [18], who has been able to obtain variances and co-variances as well as average values of mean squares.

(3) *Free random allotment (complete randomization)*. The case of no interaction with experimental units was treated in [14] (p. 73) and is discussed briefly in Section 10 of the present paper. (Other cases appear in [21], [23], [24], and [25].)

The important case of arbitrary interaction was first discussed by Wilk [16], who has treated a more complex case in [25].

Next come the relations involving a minimum of three classifications. The fact that three classifications are necessarily involved does not always appear in the corresponding analysis of variance. Thus in a randomized block experiment, *treatments*, *blocks*, and *plots* must all appear, though there is no trace of plots in the analysis of variance. The recent development of formulas for average mean squares for such relations between classifications has been in the hands of Kempthorne and Wilk, as the detailed summary will now show. The present authors have done no work in this area.

(4) *Randomized blocks*. The case of no interaction with experimental units (with plots) is classical, and formulas are well known.

The important case of interaction with experimental units was first treated, under certain restrictions, by Neyman (with cooperation of Iwaskiewicz and Kolodzieczyk) [12]. Some particular cases were followed up by McCarthy [10]. The case of arbitrary interaction (not distinguishing technical errors) was first discussed in the book of Kempthorne ([8], Sec. 8.4). This treatment was extended to the case where each treatment appears $p > 1$ times per block by Wilk [16].

(5) *Randomized fractionation (as in the classical Latin square where rows and*

columns refer to experimental units). The cases where *all* interactions are assumed zero are classical.

The case of a somewhat restricted interaction of treatments with experimental units was approached by Neyman (with cooperation of Iwaskiewicz and Kolodzieczyk) [12]. We understand that this problem is treated in [23], which we have not seen, and in "complete generality" in further unpublished work of Wilk and Kempthorne.

(6) *Crossed fractionation (as in a simple fractional factorial without blocks)*. Here the results for no interactions are classical. Nothing else seems available.

(7) *Mixed fractionation (as in a "Latin square" with two factors and one family of blocks)*. As for (6).

The work on this class of relations has been restricted to randomized blocks and Latin squares. Thus its importance depends very greatly on the field of experimentation considered.

In much of agriculture, and in many related fields, the thought of *not* having blocks within which to randomize never occurs to the statistician. The errors he faces are large, even in small blocks, and variability from experimental unit to experimental unit may dominate all other sources of variation. Blocks are all important, and confounding with blocks is common. He is almost, but not quite, justified in refusing the adjective "experimental" to a situation without blocks and in calling it "sampling" instead. But there are areas of inquiry, in parts of modern industrial technology, in the study of many measurement processes, and in other areas (usually far from biology), where errors are relatively small (compared to high-order interactions), blocks can be very large, and complete randomization is the order of the day. In these latter areas, treatment of nesting, crossing, and random allotment suffices. In the former areas, treatment of randomized blocks comes first, though it needs to be supplemented with treatments of the simpler relations. The problems selected for initial attack by the independent groups working on average values of mean squares reflect their backgrounds.

We should come now to relations involving a minimum of four classifications (for example, simple lattices). But no work seems to have been done beyond the initial classical results for no interaction.

The present paper is concerned with the crossed classifications and, in the statement of rules, with combinations of crossed, nested, and noninteracting completely randomized classifications. In its general presentation and discussion it undoubtedly makes use not only of ideas from the references and research reports cited, but also of the valuable personal discussions which its authors have had with almost all the persons mentioned above, with H. Fairfield Smith, Franklin E. Satterthwaite, the late Charles P. Winsor, and others. It would be impossible for us now to assign specific credit relating to specific ideas to specific persons. We owe a particular debt to Kempthorne and Wilk for illuminating discussions.

5. The two spans of the bridge of inference. In almost any practical situation where analytical statistics is applied, the inference from the observations to the

real conclusion has two parts, only the first of which is statistical. A genetic experiment on *Drosophila* will usually involve flies of a certain race of a certain species. The statistically based conclusions cannot extend beyond this race, yet the geneticist will usually, and often wisely, extend the conclusion to (a) the whole species, (b) all *Drosophila*, or (c) a larger group of insects. This wider extension may be implicit or explicit, but it is almost always present. If we take the simile of the bridge crossing a river by way of an island, there is a statistical span from the near bank to the island, and a subject-matter span from the island to the far bank. Both are important.

By modifying the observation program and the corresponding analysis of the data, the island may be moved nearer to or farther from the distant bank, and the statistical span may be made stronger or weaker. In doing this it is easy to forget the second span, which usually can only be strengthened by improving the science or art on which it depends. Yet a balanced understanding of, and choice among, the statistical possibilities requires constant attention to the second span. It may often be worth while to move the island nearer to the distant bank, at the cost of weakening the statistical span—particularly when the subject-matter span is weak.

In an experiment where a population of C columns was specified, and a sample of c columns was randomly selected, it is clearly possible to make analyses where

- (1) the c columns are regarded as a sample of c out of C , or
- (2) the c columns are regarded as fixed.

The question about these analyses is not their validity but their wisdom. Both analyses will have the same mean, and will estimate the effects of rows identically. Both analyses will have the same mean squares, but will estimate the accuracy of their estimated effects differently. The analyses will differ in the length of their inferences; both will be equally strong statistically. Usually it will be best to make analysis (1) where the inference is more general. Only if this analysis is entirely unrevealing on one or more points of interest are we likely to be wise in making analysis (2), whose limited inferences may be somewhat revealing.

But what if it is unreasonable to regard c columns as any sort of a fair sample from a population of C columns with $C > c$. We can (at least formally and numerically) carry out an analysis with, say, $C = \infty$. What is the logical position of such an analysis? It would seem to be much as follows: We cannot point to a specific population from which the c columns were a random sample, yet the final conclusion is certainly not to just these c columns. We are likely to be better off to move the island to the far side by introducing an unspecified population of columns "like those observed" and making the inference to the mean of this population. This will lengthen the statistical span at the price of leaving the location of the far end vague. Unless there is a known, fixed, number of reasonably possible columns, this lengthening and blurring is likely to be worth while.

This discussion follows the line of the classical discussion of "selecting the right error term," as developed by Fisher and expounded by many statisticians,

with two considerations rarely faced except in careful discussions of groups or series of experiments (cf. Chapter 28 of [8] for references):

- (1) We admit that more than a single analysis of given data may have "correctness."
- (2) We have tried to state the uncertainties of the *post-facto* $C = \infty$ choice a little more specifically than usual.

In any case, however, this discussion illustrates one way in which the nature of the appropriate analyses of variance depends on the purposes of the analysis.

6. The varied roles of randomization. Emphasis on randomization of arrangement entered modern statistics with the analysis of variance—in the early work of R. A. Fisher. The year 1935, in which Neyman (with cooperation of Iwaskiewicz and Kolodzieczyk) [12] discussed the problem of interaction with experimental units, was marked by the appearance of *The Design of Experiments*, in which Fisher stressed both the role of randomization as a guarantor of the validity of an experiment and the close correspondence, in certain examples, of tests of significance based on randomization (assuming no interaction with experimental units) with those based on an assumption of normality of distribution.

Two years later, in 1937, the first papers of Pitman's series on randomization appeared. Pitman was seeking tests of significance which would be independent of the underlying distribution, naturally tried for randomization tests, and was much surprised when the natural approximation to these tests turned out to be the classical normal theory tests. This series of papers culminated in his *Biometrika* paper [13] which dealt with randomized blocks. In the meantime, Welch [15], had applied similar methods to both randomized blocks and Latin squares. In all of these papers, the assumption was made that there were no interactions with experimental units. As clearly stated, the motivation of these papers was to obtain tests which would apply to any distribution of errors, and randomization was used to mediate this independence of distribution.

The treatment of Pitman and of Welch went far beyond the subject of the present paper, the average values of mean squares. They dealt with a function of the ratio of mean squares, and obtained a number of moments. For the cases they treated, their results go farther than our knowledge for any other situations. They treated randomized blocks and Latin squares explicitly. Implicitly their results cover any less restrictive randomization; in particular, their results also cover complete randomization. Their work was carried out for the case of one classification of treatments. Implicitly, it applies to cases of two or more treatment classifications, but only if these treatment classifications do not interact.

We have today no comparable basis for the analysis of factorial experiments where interactions may be present and where, hence, the main effects will usually be compared with interaction. Our knowledge is limited to average values of mean squares, except for the work of Hooke [18] on second moments in the two-way case. To reach a situation comparable to the case of a single classification of treatments will require an extension of Hooke's work, both to higher moments and to more-way classifications. The essential difficulty is the probable existence

and possible perversity of interactions. No assumptions about randomization will allow us to avoid facing them. For the present, presumably, we shall continue to base our inferences on average values of mean squares, and try to comfort ourselves with distant and tenuous analogies with the single treatment classification case where, if interactions with experimental units be absent, Fisher, Pitman, and Welch have shown us that the situation is rather pleasant.

If one is dealing with situations where the contributions of the experimental units to variability is large or even dominant, it is easy to alter the role of randomization somewhat. Instead of thinking of it as a mediator which assures the validity of significance tests for any shape of error distribution, as Pitman did, one can think of it in itself. (If the effects of experimental units dominate all other sources of variation, it is not only easy but necessary.) This point of view was vigorously taken up by Kempthorne [8], [9], and has strongly motivated the work of Wilk and Kempthorne [21], [22], [23], [24], [25], [26].

While we hold rather definite views, we do not feel that the issues involved have been finally settled. We do feel, however, that a knowledge of what the issues are is essential in understanding just how far average values of mean squares take us and how our work is related to that of others.

DESCRIPTION OF SITUATIONS TREATED

7. A description from the point of view of urn sampling. We consider a finite or infinite number of elements (possible measurement results on animals, plots, batches, samples, etc.). These elements are classified into R rows (litters, days, blocks, pressures, temperatures, times, etc.) and C columns (doses, operators, treatments, temperatures, times, catalysts, etc.). Each of the RC pigeonholes thus formed contains N elements. In some circumstances it is natural to consider the RCN elements as a population, in others it would be unnatural.

A sample of r rows is taken, each row having probability r/R of being selected. Given that a particular row has been selected, the conditional probability that any other given row will have been selected is $(r-1)/(R-1)$.

Similarly a sample of c columns is taken, each column having probability c/C of being selected. Given that a particular column is in the sample, the conditional probability that any other given column will be in the sample is $(c-1)/(C-1)$.

Every pigeonhole located in a sampled row and a sampled column is thus a cell included in the sample. The pattern of cells so obtained might well be called a bisample, since it is defined by two samples, one of rows and one of columns. In each such cell n elements are sampled, each element in the sampled cell having probability n/N of being selected. Given that a particular element has been selected from this cell, the conditional probability that any other element in the cell has also been selected is $(n-1)/(N-1)$.

These are the only assumptions we shall need (for the situation with which we are concerned). From the point of view of urn sampling, we have only to define the variance components corresponding to such an array of RCN elements.

Notice that our definitions of variance components apply to fixed models and mixed models as well as to random ones. There has been a tendency to refer to random models as variance components models, presumably because the analysis associated with a fully random model is most likely to be directed toward the estimation of variance components. This choice of words has already given rise to confusion in connection with mixed models, and its continued use will undoubtedly cause other complications. We recommend that it be discarded.

Since we have to deal with two sets of rows, columns, and cells, one set in the original array of *RCN* elements and one in the array of *rcn* elements we actually have at hand, it is unusually important to make a clear distinction in our notation. We shall do this by using capital letters for all that has to do with the underlying array and lower case for all that has to do with the observed array. Thus

x_{ijk} = value of k th element in j th column in the i th row in the observed array,

X_{IJK} = value of K th element in the J th column in the I th row in the underlying array.

(This convention will be altered in Sections 23ff). We shall indicate an unweighted mean over observed values by a dot in place of the subscript averaged over, and over underlying values by a dash in place of the subscript averaged over. Thus

$x_{ij\cdot}$ = mean value of all observed elements in the i th row and j th column,

X_{IJ-} = average value of all underlying elements in the pigeonhole in the I th row and J th column.

We can now define the variance components of the underlying array by

$$\sigma_R^2 = \frac{1}{R-1} \sum_{I=1}^R (X_{I--} - X_{---})^2,$$

$$\sigma_C^2 = \frac{1}{C-1} \sum_{J=1}^C (X_{-J-} - X_{---})^2,$$

$$\sigma_{RC}^2 = \frac{1}{(R-1)(C-1)} \sum_{I=1}^R \sum_{J=1}^C (X_{IJ-} - X_{I--} - X_{-J-} + X_{---})^2,$$

$$\sigma_R^2 = \frac{1}{RC} \sum \sum \sigma_{IJ}^2,$$

where

$$\sigma_{IJ}^2 = \frac{1}{N-1} \sum_{K=1}^N (X_{IJK} - X_{IJ-})^2.$$

These formulas reduce to the standard ones in all familiar special cases.

We observe that the variance components thus defined are exactly the mean squares we would obtain if all the values in the entire pigeonhole model were subjected to analysis of variance, except for factors of N , NR or NC .

The description we have just given is both less and more general than the model of Section 3. It is less general because we assumed that all populations

were of the same size. It is more general because we did not require that the sampling be purely at random, but only that any pair of rows, columns, or elements was as likely to enter the sampling as any other. We shall now see that these differences are nonessential.

We assumed a constant population size for the purely expository reason that we had not introduced the subscripts when we mentioned population size. We can now alter three sentences or clauses to read: "The IJ th pigeonhole contains N_{IJ} elements," "each element in the sampled cell having probability n/N_{IJ} of being selected," "the conditional probability ... is $(n-1)/(N_{IJ}-1)$." After this alteration the urn sampling model is at least as general as our initial model (see Section 3).

In the initial model we assumed that all samplings were "at random." Since we are only concerned with averages of quadratic functions of the sampled elements, it would be merely an application of a general principle to conclude that if the results hold for sampling "at random," they must also hold for sampling "in which *individuals* and *pairs* have equal chances of being selected." We may, if we wish, consider that the initial model has been so generalized. The initial model and the urn sampling model then became equivalent.

8. Linear models with "tied" interaction. For convenience in exposition, we shall take $n = 1$ and drop the index k for most of this section. Under these conditions, the conventional linear model of most texts would appear like this:

$$x_{ij} = \theta + \xi_i + \eta_j + \omega_{ij},$$

where the ξ_i are the row contributions (sometimes called "effects"), the η_j are the column contributions, and the ω_{ij} are error or discrepance contributions. Various normalizing conditions may or may not be applied. The ξ 's, the η 's, and the ω 's will be variously assumed to be fixed or random samples from infinite (or perhaps finite) populations. But the key assumption will be that the variation of the ω 's is *independent* of what the ξ 's and η 's may be. It is this assumption of independence which has made the use of such linear models so special and dangerous.

In particular, this model cannot accommodate the following situation easily described in terms of four pigeonholes:

$N(0, 1)$	$N(2, 1)$
$N(2, 1)$	$N(0, 1)$

where $N(\mu, \sigma^2)$ stands for a normally distributed infinite population with average value μ and variance σ^2 . The independence assumption is an assumption about the behavior of the world, and not just about how we do experiments.

Although some have expressed doubts, there are real advantages to linear models.

It has therefore been worth while to learn how to generalize linear models to apply without assumption.

Let us go, therefore, to a situation with R potential rows and C potential columns, where I designates a potential row and J designates a potential column. If $f(I, J)$ is an arbitrary function of I and J , we may choose to define

$$\begin{aligned}\theta &= f(-, -), \\ \xi_I &= f(I, -) - f(-, -), \\ \eta_J &= f(-, J) - f(-, -), \\ \lambda_{IJ} &= f(I, J) - f(I, -) - f(-, J) + f(-, -),\end{aligned}$$

where replacement of an " I " by a " $-$ " implies averaging over all R potential rows and replacement of a " J " by a " $-$ " implies averaging over all C potential columns. (Note that other definitions are possible.)

We shall then have

$$f(I, J) = \theta + \xi_I + \eta_J + \lambda_{IJ},$$

where

$$\lambda_{I-} = \text{a constant} = \lambda_{-J}$$

(and where indeed this constant is zero, and $\sum \xi_I = 0$ and $\sum \eta_J = 0$, although we shall *not* require any of these to vanish when we use this model).

Now if we pick r rows out of R and c columns out of C , we may designate actual rows and columns with i 's and j 's, and write $I(i)$ for the potential row corresponding to actual row i , and $J(j)$ for the potential column corresponding to the actual column j . If x_{ij} is the value of $f[I(i), J(j)]$, we have

$$x_{ij} = \theta + \xi_i + \eta_j + \lambda_{ij},$$

where we have written ξ_i for $\xi_{I(i)}$, η_j for $\eta_{J(j)}$, and λ_{ij} for $\lambda_{I(i)J(j)}$. We have here an additive model for a bisample of rows and columns drawn from an arbitrary $R \times C$ array of constants. How does this differ from the "independent" model in which ω_{ij} is assumed independent of ξ_i and η_j ? How is it to be described?

It differs in that the λ_{ij} are "tied" to the corresponding ξ_i and η_j . They are "tied" in a very specific way, however. It need not be true that λ_{ij} is a well-defined function of the values of ξ_i and η_j , for there may be, for example, various pairs of values of I and J for which the corresponding ξ_I and η_J have the same values but the corresponding λ_{IJ} are not equal. Thus λ_{ij} is tied to ξ_i and η_j through the values of $I(i)$ and $J(j)$ rather than through the values of ξ_i and η_j .

But what if the pigeonholes contain populations, and we sample one element

from each actual cell. Can this be treated similarly? Quite easily. Let

$$f(I, J) = X_{IJ-} = \frac{1}{N_{IJ}} \sum_K X_{IJK},$$

$$\omega_{IJK} = x_{IJK} - f(I, J).$$

Then, if the sampling in each cell is at random and independent, ω_{ij} is a random sample of 1 from a population $D_{I(i), J(j)}$ whose average value is zero, and

$$\begin{aligned} x_{ij} &= X_{I(i), J(j), K} = f(I, J) + \omega_{ij} \\ &= \theta + \xi_i + \eta_j + \lambda_{ij} + \omega_{ij}, \end{aligned}$$

where, although λ_{ij} and ω_{ij} have the same indices, they represent quite different sorts of quantities. The λ 's are interactions tied to the ξ 's and the η 's, while the ω 's are independent fluctuations.

9. Main effects and main contributions. All those who use the analysis of variance are familiar with the words "main effect," but far fewer have any really clear understanding of what they mean. Yet all specific analysis of variances procedures imply very specific interpretations of what it is that concerns us. Any practically satisfactory structure for the analysis of variance must bring the essential definitions into the limelight.

Because there is likelihood of confusing the quantity in the model and its estimate derived from the observations, we call the quantity in the model a *main contribution*. This is its name, but what is it? Usually only a relative definition makes sense. We talk of main contributions, but we work with their differences. What then does the difference $\xi_i - \xi_k$ mean in a situation modeled by

$$x_{ij} = \theta + \xi_i + \eta_j + \lambda_{ij} + \omega_{ij}$$

with the two conditions

$$\begin{aligned} \lambda_{I-} &= \text{a constant} = \lambda_{-I}, \\ \text{ave } \{\omega_{IJ}\} &= 0. \end{aligned}$$

Let, again,

$$f(I, J) = \text{average response in the } IJ \text{ pigeonhole};$$

then

$$\theta + \xi_i + \eta_j + \lambda_{ij} = f(I, J),$$

and we easily find that

$$\xi_i - \xi_k = f(I, -) - f(K, -) = \frac{1}{C} \sum [f(I, J) - f(K, J)].$$

This states that the difference in main contributions between the I th and K th rows is the average over all C potential columns of the average effect of changing

from the I th to the K th row. The average is over *all* potential columns, and for the present definition is *equally* weighted. What is most important in defining row main contributions is our choice of what are the potential columns.

This is the heart of the problem of main contributions—it is our attitude toward the *other* factors which affects the definition of a main contribution. A change from considering *rows* as fixed to considering rows as a sample from a large population need have no effect at all on the definition of *row* main contributions, but it will substantially alter the definitions of the *column* main contributions. It is only when we have relatively explicit definition that we can force ourselves to recognize this fact.

We stated that an alteration in the set of potential rows *need* not alter the definition of the row contributions. This is so because we have *not* required that $\sum \xi_i = 0$. We have not made this requirement because it serves no useful purpose. (Its imposition would make all the ξ_i estimable from an experiment, but there are two reasons why this would not be useful: (i) Because we are interested in estimating only the differences $\xi_i - \xi_j$ anyway. (ii) Because, usually, only some of the ξ_i appear as $\xi_i = \xi_{I(i)}$, and since the others can surely not be estimated, what do we care about one more unestimable parameter? By making one more parameter unestimable, we have gained inestimable freedom.)

It is in presenting explicit quantities which represent main contributions that this linear model—the generalized, nonindependent linear model—has a significant and specific role. It contains exactly the same definition of main contributions, but how many readers recognized this as they read this section? It is concealed in such urn formulas as

$$\sigma_R^2 = \frac{1}{R-1} \sum_{i=1}^R (X_{I--} - X_{---})^2,$$

which implicitly state that what we shall ask the "row" line of the analysis of variance to inform us about, are the X_{I--} . The difference $X_{I--} - X_{---}$ is identical with $\xi_I - \xi_K$, but the linear model brings the situation out with greater force and clarity.

10. Description by linear models. We now set out the general linear model in the form

$$x_{ijk} = \theta + \xi_i + \eta_j + \lambda_{ij} + \omega_{ijk},$$

where $1 \leq i \leq c$, $1 \leq j \leq r$, $1 \leq k \leq n$ and the assumptions are as below, where in dealing with simple finite populations, we shall always use the modern definition of variance, dividing by *one less* than the number of elements involved. Thus, for example, the variance of the population of potential columns is

$$\sigma_\eta^2 = \frac{1}{C-1} \sum_{j=1}^C (\eta_j - \bar{\eta})^2,$$

where η_- is the average of η_J over the population of potential columns

$$\eta_- = \frac{1}{C} \sum \eta_J.$$

In general, we shall use dashes for averages of population quantities, and dots for means of sampled quantities.

With these preliminaries we can state the assumptions about θ , ξ_i , and the η_j which we wish to apply:

- (1) There is a population of general contributions, of variance σ_θ^2 , and θ is a random sample of 1 from this population.
- (2) There is a population of row contributions of size R and variance σ_R^2 and $\xi_1, \xi_2, \dots, \xi_r$ are a random sample of size r from this population.
- (3) There is a population of column contributions of size C and variance σ_C^2 , and $\eta_1, \eta_2, \dots, \eta_c$ are a random sample of size c from this population.
- (4) There is a two-way array of interactions λ_{IJ} , one for each row contribution and column contribution in the corresponding populations, the averages (over J) λ_{I-} are independent of I , the averages (over I) λ_{-J} are independent of J , and the interaction actually occurring in the expression for x_{ij} is $\lambda_{I(i),J(j)}$; that is to say, it is the interaction which corresponds to the column contribution and row contribution which occur in the expression for x_{ij} .
- (5) The sampling in (1), (2), and (3) takes place independently.

We define, for reference,

$$\begin{aligned} \sigma_{RC}^2 &= \frac{1}{(R-1)(C-1)} \sum \sum (\lambda_{IJ} - \lambda_{-})^2 \\ &= \frac{1}{(R-1)(C-1)} \sum \sum (\lambda_{IJ} - \lambda_{I-} - \lambda_{-J} + \lambda_{--})^2. \end{aligned}$$

(Note that we do *not* use $(RC-1)$ as a divisor.)

There remain the assumptions about the ω_{ijk} , where we still have much choice, so long as we require that

$$\text{ave } \{\omega_{IJK}\} = \text{a constant independent of } I \text{ and } J.$$

The assumption which most exactly corresponds to the pigeonhole model is the following:

- (6') For each of the RC combinations of a population row with a population column there is a population of size N_{IJ} , average value μ (the same for all IJ), and variance σ_{IJ}^2 . The ω_{ijk} are a sample of n from the $I(i)J(j)$ th population. Sampling is at random in each cell and independent, both between cells and of the samplings in (1), (2), and (3).

With this choice, the generalized linear model corresponds exactly to the pigeonhole model, provided we place $\sigma_\theta^2 = 0$ and thus keep the general contribution constant.

The variance component for repetition (or "error") σ_R^2 , and the effective population size N , are to be found from

$$\sigma_R^2 = \frac{1}{RC} \sum_I \sum_J \sigma_{IJ}^2.$$

$$\left(1 - \frac{n}{N}\right) \sigma_R^2 = \frac{1}{RC} \sum_I \sum_J \left(1 - \frac{n}{N_{IJ}}\right) \sigma_{IJ}^2.$$

We remarked that (6') was not the only choice for a linear model. There are many. We shall give *detailed* consideration to only one other, namely:

(6'') The cell contributions are

$$\omega_{ijk} = \omega'_{ijk} + \omega''_{ij},$$

where the ω'_{ijk} satisfy the conditions of (6') while the ω''_{ij} are the result of independently randomizing a set of rc values of variance σ^2 among the rc actual cells.

We shall now need to set

$$\sigma_R^2 = \frac{1}{RC} \sum_I \sum_J \sigma_{IJ}^2$$

$$\left(1 - \frac{n}{N}\right) \sigma_R^2 = \sigma^2 + \frac{1}{RC} \sum_I \sum_J \left(1 - \frac{n}{N_{IJ}}\right) \sigma_{IJ}^2$$

(the last of which may lead to negative N 's).

Clearly (6'') is more complicated than (6'), which matched the pigeonhole model. Why then do we wish to consider this added complexity? Because experiments are often more complicated than the simple pigeonhole model makes allowance for. For example, let us suppose that we wish to study a chemical reaction at all combinations of 6 specific pressures and 7 specific temperatures. The pigeonhole model would naturally have 6 rows and 7 columns, and in each pigeonhole we would put an infinite population. The pattern of the averages of these populations would represent the response of this process to pressure and temperature. The variations within each of these populations would represent fluctuations in process behavior and measurement. But we would be ill advised to stop here. Every well-trained statistician would insist, especially if every combination of pressure were to be tried once only, that the order in which the experiments were performed should be randomized. He would do this in fear of systematic errors somehow associated with time. In other words, in fear that the ω''_{ij} of (6'') were not all zero. We need very great flexibility in our models to deal with real situations such as this. (Notice that we do not attempt to discuss the very real and important cases where randomized blocks, and related uses of randomization are involved. We are here concerned with the flexibility required for the simple case of two crossed classifications, where experimental units are not so important as to make complete randomization inadequate.)

The flexibility of the linear model is not limited to (6"). It can easily accommodate any number of terms of each or all of the following kinds:

- (a) Individual distributions of ω'_{IJK} for each IJ .
- (b) Randomization of rc values ω'' over the rc cells.
- (c) Randomization of a sample of rc values ω''' drawn randomly from a population of size $M > rc$ over the rc cells.
- (d) Randomization of RC values ω'''' over the RC pigeonholes.
- (e) Randomization of RC values ω'''' drawn randomly from a population of size $M > RC$ over the RC pigeonholes.

The ease with which such complete randomized contributions can be added to linear models depends on a simple remark, which was apparently first made by McCarthy ([10], p. 358) in the case where all is normal and the variances are constant and was later exploited ([14], p. 107) in the general case; namely, that equal, completely symmetric correlations do not affect the average mean squares, except for the mean square for the general mean (to be discussed in Section 14). That this is true can be easily seen by the following argument. If the correlations are equal and negative, add a varying general contribution with variance equal to the correlations. This will exactly annul the correlations. If the correlations are equal and positive, observe that they are exactly the correlations which would result from such a fluctuating main effect. In either case, the desired result follows immediately.

It may not be immediately appreciated why we need more than one term of a given kind. However, if, in the chemical reaction experiment just considered, we plan to use a separate piece of equipment and a separate operator for each run, we will wish to randomize all three variables—epoch, equipment, and operator. Our model will require three terms of the ω'' kind. There will be three different sorts of experimental units!

If we put down all the variables considered in a really complex experiment, even the linear model begins to look complicated.

11. The development of the linear model. The classical linear model, as described at the beginning of the previous section, had the following disadvantages from our present point of view.

- (1) It made assumptions about the way in which the response depended on the two (or more) factors.
- (2) It assumed that at most one term should appear for a given set of indices.
- (3) It assumed special normalizations (like $\lambda_{I-} = 0$) rather than more general ones (like λ_{I-} independent of I).
- (4) It assumed constant variance "within pigeonholes."

(These have been listed in general order of importance.)

Models which avoided the first disadvantage were introduced (for randomized blocks and Latin squares) by Neyman (with the cooperation of Iwaskiewicz and Kolodzieczyk) [12] and were used by McCarthy [10], but, perhaps because of the generally negative flavor of the papers, seem to have lain forgotten for

many years. The results (for crossed classifications) of Memorandum Report 18 [21] were expounded in the terms used in the last section. Similarly models were independently and extensively used (for various situations) in the book of Kempthorne [8] and have been prominent in the recent work of Wilk and Kempthorne. For all of this, the relationship of λ_i to ξ_i and ξ_j when it is tied through the values of $I(i)$ and $J(j)$ is not yet as widely familiar among statisticians, as it seems to us that it should be.

The introduction of two or more terms with the same set of subscripts was also due to Neyman (with the co-operation of Iwaskiewicz and Kolodzieczyk) [12] who introduced first a single term and then separated it into two parts, corresponding to "plot error" and "technical error." In [14], the approach was to start from the parts and then combine, rather than the reverse.

Both this difference, and the introduction of the weaker normalization in [14], were related to two desires: a desire to weaken assumptions wherever possible, and a desire to treat contributions more as things with independent existence rather than as differences between certain averages. The general philosophy ran along the lines that "if the effect is real and substantial, it should appear in the model whether or not it can be estimated from the data."

The assumption of constant variance "within pigeonholes" is a natural assumption, and is important in connection with both higher moments of mean squares and with the variances of contrasts and other linear combinations. As pointed out in [14], however, it may be dropped without any effect on average mean squares.

A more recent development in the case of linear models has been the introduction of a pair of closely related linear models, called "the population model" and "the statistical model" by Wilk [22] and their further use by Wilk and Kempthorne [23], [24]. By introducing this distinction, the assumptions about the response to the various classifications can be presented first, while the assumption about the randomization involved in setting up the experiment can be added later. This development formalizes and makes definite distinctions hinted at in Kempthorne's book [8].

12. The next consideration. The next step in generality, when we have completely randomized some contribution, to which we will attach the name "experimental units," is to consider the possibility that these units can interact with the other classifications. (Since the results are a function of both classifications jointly, the behavior resembles that of crossed classifications; yet, in any particular experiment, one classification is nested in the other. For these reasons, the term "cross nested" has been used informally to describe the situation.)

It is easiest to make the situation clear when a one-way classification, say temperature, is involved, and epochs are randomized. If we admit that the effect of an epoch can be different for different temperatures, then we are led to a two-way set of pigeonholes—pigeonholes labelled by temperature and epoch. If each temperature is used once only, the actual experiment (if fully randomized) could be described as picking a set of pigeonholes, one in each row and column,

and picking a value from each of these selected pigeonholes. It is as if we had the results of that part of a Latin square occupied by a single letter. Even though we cannot estimate the temperature by epoch interaction variance component, it is clear that it will enter into the average values of the mean squares which we do obtain.

Complete randomization is a binary relation between classifications, but, especially when without replication, it is closely related to the ternary relation of the Latin square. Again independent work has led to related results. Kempthorne treated the Latin square with arbitrary interactions between rows and columns, but with no interactions with treatments in his book ([8], pp. 190-191). In particular he obtained the variance of a treatment mean. This last particular result is also the variance of the general mean in the unreplicated case of complete randomization with arbitrary interaction (with experimental units) which was obtained independently by Cornfield and Evans, and reported, with a modified proof, in Hansen, Hurwitz, and Madow ([6], pp. 262-265). So far as complete randomization is concerned, more general results were obtained by Wilk [16, 25].

It is clear, however, that there are a number of stages of sophistication, care, or cynicism (as you please) about our treatment of such a factor as epoch. Some of these are the following:

- (1) In both design and analysis we neglect it altogether.
- (2) We neglect it in design, and in analysis we use the additive model to show ourselves that we should have randomized it after all.
- (3) We randomize it in design, and in analysis we use the additive model to show ourselves how well off we are.
- (4) We randomize it in design, and in analysis we use the arbitrary-interaction model to show ourselves that the situation is not quite perfect.
- (5) We take it into the design of the experiment as a full-fledged factor.

Each of these attitudes is appropriate in its place. In every experiment there are many variables which *could* enter, and one of the great skills of the experimenter lies in leaving out only inessential ones. What we have just observed is that there are gradations between variables which are entirely out and those which are entirely in.

It is generally understood that in the design of an experiment, there are three classes of variables:

- (i) those treated as factors,
- (ii) those randomized, and
- (iii) those neglected.

The additive model helps us to think about the choice between neglect and randomization. The arbitrary-interaction model helps us to think about the choice between randomization and recognition as a factor.

RESULTS

13. Results for the two-way classification. The average mean squares for the pigeonhole model are set forth for the general case and for three special cases

generalizing the usual fixed, random, and mixed models in Table 1. The important thing to notice in this table is the appearance of the factors

$$1 - \frac{c}{C}, \quad 1 - \frac{r}{R}, \quad 1 - \frac{n}{N},$$

which serve to suppress certain terms completely in the fixed situation and to reduce their effects on the average values of the mean squares when the populations are finite but larger than the samples.

TABLE 1
General results and special cases

Line in the analysis of variance	Average value of mean squares			
	General case	Special cases		
		(Fixed) $c = C, r = R$ N infinite	(Random) c, r finite N, C, R , infinite	(Mixed) $c = C, r$ finite N, R infinite
Rows	$\left(1 - \frac{n}{N}\right) \sigma_E^2$ $+ \left(1 - \frac{c}{C}\right) n \sigma_{RC}^2$ $+ n n c \sigma_R^2$	$\sigma_E^2 + n n c \sigma_R^2$	$\sigma_E^2 + n \sigma_{RC}^2$ $+ n n c \sigma_R^2$	$\sigma_E^2 + n n c \sigma_R^2$
Columns	$\left(1 - \frac{n}{N}\right) \sigma_E^2$ $+ \left(1 - \frac{r}{R}\right) n \sigma_{RC}^2$ $+ n r \sigma_C^2$	$\sigma_E^2 + n r \sigma_C^2$	$\sigma_E^2 + n \sigma_{RC}^2 + n r \sigma_C^2$	$\sigma_E^2 + n \sigma_{RC}^2 + n r \sigma_C^2$
Interaction	$\left(1 - \frac{n}{N}\right) \sigma_E^2 + n \sigma_{RC}^2$	$\sigma_E^2 + n \sigma_{RC}^2$	$\sigma_E^2 + n \sigma_{RC}^2$	$\sigma_E^2 + n \sigma_{RC}^2$
Error	σ_E^2	σ_E^2	σ_E^2	σ_E^2

Notice in particular how, when $R = C = N = \text{infinity}$, σ_E^2 follows σ_{RC}^2 everywhere, just as the lamb followed Mary. This is the usual result for a random model, and appears here in a situation of far greater generality.

The disappearance of σ_{RC}^2 from the average values of row and column mean squares in the fixed case is also familiar.

In the mixed model the average mean square for rows does not involve the interaction variance component (because all columns were observed!), while the average mean square for column involves it with unit weight (because only a small sample of rows were observed!). In each case the behavior of the *other variable* (or, in general, the *other variables*) determines whether the interaction appears or not.

That the result for the mixed model is a consequence solely of the urn sampling approach, and not of the special definitions of interaction and row components of variance used, may be seen from the following simple example. Consider a population composed of three rows and two columns, with one element per cell, having the following numerical values:

0	100
100	0
50	50

Each of the three possible samples of two rows and two columns will yield a row mean square of zero and an interaction mean square greater than zero. Hence for this population the interaction mean square exceeds the row mean square for each sample. Testing the row mean square against interaction would be obviously incorrect.

These results are very general. We assumed only a symmetrical placing of *RC* sets of numbers in *RC* pigeonholes and a well-defined method of extracting the *rcn* numbers entering the actual analysis of variance. Otherwise, the numbers can be as arbitrary as we please.

14. The mean square for the general mean. While the central facts have been pointed out from time to time (e.g., Cochran [3], Hendricks [7]), and while a number of mathematical statisticians have made use of the fact in their lectures, it does not seem to be widely recognized that it is almost always feasible to complete the analysis of variance table by adding a line for the general mean, and that when this has been done, both classical analysis of variance and sampling from finite populations become special cases of a unified situation. We feel that wider recognition and use of this fact would be valuable.

In the present situation we have only to introduce a reference origin *M* (which is freely at our disposal) and define both the sum of squares for the general mean and the mean square for the general mean (there is but one degree of freedom) as

$$rcn (x... - M)^2$$

(This is what is frequently known as the "correction term," and is appropriate to the use of *M* as a computing origin.) It is easy to show that we have the following average value of the mean square for the mean

$$rcn\sigma_r^2 + cn \left(1 - \frac{r}{R}\right) \sigma_R^2 + rn \left(1 - \frac{c}{C}\right) \sigma_C^2 + n \left(1 - \frac{c}{C}\right) \left(\frac{1-r}{R}\right) \sigma_{RC}^2 \\ + \left(1 - \frac{n}{N}\right) \sigma_x^2 + rcn(\text{ave } \{x...\} - M)^2,$$

which like all average mean squares in a balanced analysis of variance decompose into $(F) + (G) + (H)$ where

(F) = number of units involved in each corresponding mean, (rcn in the example),

(G) = variance (so far as differences are concerned) of the corresponding mean,

$$\sigma_c^2 + \left(\frac{1}{r} + \frac{1}{R}\right) \sigma_R^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_c^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{RC} \\ + \frac{1}{rc} \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_n^2 \text{ in the example,}$$

(H) = variance of corresponding contributions (usually measured among themselves, but in this case necessarily measured from the arbitrary origin, M , since there is only one general mean and hence = $(\text{ave } \{x_{...}\} - M)^2$).

15. Results for the three-way classification. The pigeonhole model for the three-way classification is the natural generalization of the two-way pigeonhole model. There are RCS pigeonholes, one in each combination of R rows, C columns, and S slices. In each pigeonhole there is a population. Random samples of r rows from R , c columns from C , and s slices from S are independently drawn. The intersections of the drawn rows, columns, and slices determine rcs cells. In each of these rcs cells n elements are drawn at random. All drawings, whether of rows, columns, slices, or elements are independent.

The variance components are defined in a way similar to that for the two-way. They differ from the (hypothetical) mean squares we would get, if we analyzed the entire model, by simple factors such as r , c , s , etc.

The average values of the mean squares are given for the general case in Table 2. Here, with obvious extensions of notation,

$$\sigma_n^2 = \frac{1}{RCS} \sum_i \sum_j \sum_k \sigma_{ijk}^2, \\ \left(1 - \frac{n}{N}\right) \sigma_R^2 = \frac{1}{RCS} \sum_i \sum_j \sum_k \left(1 - \frac{n}{N_{ijk}}\right) \sigma_{ijk}^2,$$

and

$$\sigma_G^2 = \text{variance of the general contribution.}$$

(Except for the top line, these results will be found in Bennett and Franklin ([2], p. 394), while generalizations have been given by Kempthorne and Wilk [23] and by Wilk [25]).

While the results for the three-way are not too complex, and are obviously systematic and orderly, they do take up considerable space. If we are to set out the corresponding results for factorial arrangements with more factors, there will be a premium on more compactness.

16. Unreplicated factorials in general. We can obtain this compactness by setting up some rules which will give the coefficient of any variance component

TABLE 2

Average values of mean squares in the general (replicated) three-way classification

Item	DF	Average value of mean square
General mean	1	$\begin{aligned} &\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{r}{R}\right)\left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 \\ &+ nr\left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{CS}^2 + nc\left(1 - \frac{r}{R}\right) \\ &\cdot \left(1 - \frac{s}{S}\right)\sigma_{RS}^2 + ns\left(1 - \frac{r}{R}\right)\left(1 - \frac{c}{C}\right)\sigma_{RC}^2 \\ &+ ncr\left(1 - \frac{s}{S}\right)\sigma_B^2 + ncs\left(1 - \frac{r}{R}\right)\sigma_R^2 \\ &+ nrs\left(1 - \frac{c}{C}\right)\sigma_C^2 + nc rs \sigma_g^2 \\ &+ nc rs (\text{ave } \{x\} - M) \end{aligned}$
Rows (<i>R</i>)	$r - 1$	$\begin{aligned} &\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 \\ &+ nc\left(1 - \frac{s}{S}\right)\sigma_{RS}^2 + ns\left(1 - \frac{c}{C}\right)\sigma_{CR}^2 + nc s \sigma_R^2 \end{aligned}$
Columns (<i>C</i>)	$c - 1$	$\begin{aligned} &\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{r}{R}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 \\ &+ nr\left(1 - \frac{s}{S}\right)\sigma_{CS}^2 + ns\left(1 - \frac{r}{R}\right)\sigma_{CR}^2 + nr s \sigma_C^2 \end{aligned}$
Slices (<i>S</i>)	$s - 1$	$\begin{aligned} &\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{r}{R}\right)\left(1 - \frac{c}{C}\right)\sigma_{RCS}^2 \\ &+ nr\left(1 - \frac{c}{C}\right)\sigma_{CS}^2 + nc\left(1 - \frac{r}{R}\right)\sigma_{RS}^2 + nc r \sigma_S^2 \end{aligned}$
<i>RC</i>	$(r - 1)(c - 1)$	$\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 + ns\sigma_{RC}^2$
<i>RS</i>	$(r - 1)(s - 1)$	$\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{c}{C}\right)\sigma_{RCS}^2 + nc\sigma_{RS}^2$
<i>CS</i>	$(c - 1)(s - 1)$	$\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\left(1 - \frac{r}{R}\right)\sigma_{RCS}^2 + nr\sigma_{CS}^2$
<i>RCS</i>	$(r - 1)(c - 1)(s - 1)$	$\left(1 - \frac{n}{N}\right)\sigma_B^2 + n\sigma_{RCS}^2$
Repetition	$rcs(n - 1)$	σ_B^2

in the average value of any mean square

We begin with the case without replication and the almost obvious

Rule 1. Unless the indices of the variance component include all those of the mean square, the coefficient vanishes.

As a result of this rule, we can confine our attention to cases where the indices fall into three categories:

- (1) those appearing in both mean square and variance component,
- (2) those appearing in variance component but not in mean square,
- (3) those appearing in neither.

In these terms we can now state

Rule 2. Any coefficient which does not vanish because of Rule 1 is the product of a small letter for each index which fails to appear in the variance component and of a factor

$$1 - \frac{(\text{small letter})}{(\text{capital letter})}$$

for each index in the variance component which does not appear in the mean square.

Let us discuss some examples. Consider the coefficient of σ_{cs}^2 in the mean square for columns in an unreplicated three-way. Here the behavior of the indices is as follows:

- (1) *C* appears in both,
- (2) *S* appears only in the variance component,
- (3) *R* appears in neither.

The coefficient is, therefore,

$$r \left(1 - \frac{s}{S} \right)$$

and, when we recall that $n = 1$, we see that this checks the entry in Table 2.

Consider a seven-way classification with indices *R, C, D, E, F, G, S* and the average value of the mean square for the *RDG*-interaction. What are the coefficients of $\sigma_{RCD\mathcal{E}F}^2$ and $\sigma_{RCD\mathcal{E}GS}^2$? Since *G* is not an index of $\sigma_{RCD\mathcal{E}F}^2$, its coefficient is zero by rule 1. For the second variance component, the indices behave as follows:

- (1) *R, D*, and *G* appear in both,
- (2) *C, E* and *S* appear only in the variance component,
- (3) *F* appears in neither.

By rule 2, the coefficient is

$$f \left(1 - \frac{c}{C} \right) \left(1 - \frac{e}{E} \right) \left(1 - \frac{s}{S} \right).$$

The rules are relatively easy to apply.

We can summarize the action of both rules in a 2×2 table to be applied once for each subscript. This is done for a general subscript *Q* in Table 3.

17. More general designs. Two remarks will conspire to lead us to more general expressions of the rules. First, we note a remarkable similarity between these coefficients, and those which arise when one classification is nested within another. Second, we notice that there is a useful sense in which a replicated factorial is not purely factorial in structure—in the pigeonhole model we allowed a

TABLE 3

2×2 table giving the factor in the coefficient due to any one subscript, and thus expressing both rules for factorials

Subscripts of mean square	Subscripts of variance component	
	Q present	Q absent
Q present.....	1	0
Q absent.....	$1 - \frac{q}{Q}$	q

population to nest in each pigeonhole. We are impelled to seek some general rules which apply when classifications are crossed (as in a factorial) or nested in any way. (We shall find that noninteracting completely randomized contributions are more or less automatically included.)

These relations are not the only important relations between categories. We have mentioned randomized blocks above. There are also, for example, the relations involved in balanced and unbalanced incomplete blocks, to which our rules will not apply. But crossing and nesting are the simplest, and a treatment for any combination of them will be well worthwhile.

The scope of such a treatment will be broader than just the arrangements so far mentioned. It is quite possible, for example, to have a one-way array of small pigeonholes nested in each of a two-way array of large pigeonholes and a population nested in each of the small pigeonholes. This will occur in the example of a chemical reaction where temperature and pressure define the large pigeonholes, for instance, if we repeat the reaction several times for each pressure and temperature combination and analyze several portions of each result. Here "sampling and analysis" requires a population nested in "batch," while a one-way array of "batches" is nested in every "temperature-pressure" pigeonhole.

The general rules are based on a system of indices such that a mean square or variance component referring to something of smaller scope must have all the indices of the quantity of larger scope, and one or more in addition. A possible set of indices for the modified chemical reaction example just described might be

T = temperature,

P = pressure,

PT = their interactions,

BPT = batches (within PT combinations).

$EBPT$ = sampling and analysis (within batches).

In this example it makes no sense to define a B main contribution across temperature or pressure, and hence we should not in this notation try to define a B mean square or a B variance component. Indeed we should also not try to use BT - or

BP-quantities of any sort. In order to inquire, for example, how many batches to use to obtain a given accuracy, we should be concerned with "batches within a temperature and pressure combination" and in this notation this is a *BPT* effect. (It would perhaps be clearer to use *B(PT)* or $B \subset PT$, but we shall not do this here.)

Any set of indices has certain indices (there may be none) toward which it behaves like an interaction. This means that if we sum the corresponding contributions over such an index (covering the pigeonhole model) the total (and hence the mean) is a constant. In the example just discussed

Set of indices	Interactionlike indices
<i>T</i>	<i>T</i>
<i>P</i>	<i>P</i>
<i>PT</i>	<i>T, P</i>
<i>BPT</i>	<i>B</i>
<i>EBPT</i>	<i>E</i>

The appearance of *B* and *E* as interactionlike indices merely means that the average contributions for "batches" and "sampling and analysis" are defined to be zero. The absence of *T* in the last two cases reflects the fact that summing over one batch at each temperature, or summing over one sample and analysis at each temperature, will not ensure the disappearance of the mean batch contribution or the mean sampling and analysis contribution.

We continue to use capital letters for the dimensions of the pigeonhole model, and small letters for the corresponding dimensions of the experiment. We can now state a single unified rule (covering crossed and nested relations, and classifications completely randomized upon them) as follows:

Rule: Divide the indices into five groups as follows:

- (1) those which appear in the mean square but not in the variance component;
- (2) those which appear in the variance component and not in the mean square and are interactionlike for the set of indices defining the variance component;
- (3) those which appear in the variance component but not in the mean square, and are *not* interactionlike;
- (4) those which appear in both;
- (5) those which appear in neither.

The coefficient with which the variance component appears in the average value of the mean square is the product of a factor for each index, the factors being as follows:

- (a) for group (1) each factor is zero,
- (b) for group (2) each factor is

$$\left(1 - \frac{(\text{small letter})}{(\text{capital letter})}\right),$$

(c) for groups (3) and (4) each factor is unity,

(d) for group (5) each factor is the corresponding small letter.

This rule applies so long as all categories are related by crossing or nesting.

In this form, this rule includes the rules of the previous section, and can also be expressed in a simple table, as in Table 4.

TABLE 4

Factor due to any subscript, Q, in the coefficient of any variance component in any average mean square for any combination of crossing and nesting

Subscripts of mean square	Subscripts of variance component		
	Q present and interactionlike	Q present, not interactionlike	Q absent
Q present	1	1	0
Q absent	$1 - \frac{q}{Q}$	1	q

18. Reasons for the general rules. As indicated by the discussion of Section 23, below, the basic reasons for the general rules are two:

(1) Mean squares are ordinarily expressed on a per-element basis, so the number of elements associated with specific values of the indices of a mean square arises as a factor.

(2) Where an index is interactionlike we are dealing with variances of sample means, so that factors of

$$\frac{1}{(\text{small letter})} \text{ minus } \frac{1}{(\text{capital letter})}$$

arise from the formula for the variance of a mean from a finite population. These reasons are equally compelling for crossing and nesting.

19. Non-interacting, completely randomized terms in linear models. We discussed briefly, at the end of Section 10, some five kinds of added terms which might be added to the linear model for the two-way classification. We summarize in Table 5 the coefficients with which the corresponding variances or mean variances will appear. These results are easily obtained by the usual way of dealing with independent linear models. Observe that if we square the model, the average value of all cross-terms will cancel out (except in the mean square for the general mean) and hence that average mean squares can be evaluated separately for the original contributions and those of these kinds.

It was emphasized (at the end of Section 10) that the easy addition of such terms was a consequence of a general principle. Several such terms can be added. If the modified chemical reaction example discussed in the last section involved two raw materials, we might have completely randomized packages of one over batches, and have made up one solution of the other for each pressure-temperature combination, randomizing the allotment of solutions. If, in addition, some

TABLE 5

Kind	Average mean of square value for rows, columns, interaction	General Mean
(a)	$1 - \frac{n}{N}$	$1 - \frac{n}{N}$
(b)	1	0
(c)	1	$1 - \frac{rc}{M}$
(d)	1	$1 - \frac{rc}{RC}$
(e)	1	$1 - \frac{rc}{M}$

other relevant item were completely randomized over individual samples, we should have a model made up of at least four parts:

- (1) dependence of yield on pressure and temperature combined with routine fluctuations and errors,
- (2) effects of first raw material,
- (3) effects of second raw material,
- (4) effects of other relevant factors.

If we can properly assume an absence of interaction between these four parts, then we will obtain four sets of (partial) variance components as follows

- (1) appears in *EBPT*, *BPT*, *PT*, *P* and *T*,
- (2) appears in *BPT*, *PT*, *P* and *T*,
- (3) appears in *PT*, *P* and *T*,
- (4) appears in *EBPT*, *BPT*, *PT*, *P* and *T*.

Because of the no-interaction-across-parts assumption, the four sets will behave entirely independently, and the rules of the last section will apply to each separately.

It will only be necessary to remember that (so long as we are not concerned with the general mean) *no* index is interactionlike for contributions which are completely randomized. This same principle will apply to other examples where absence of interaction between completely randomized parts is appropriately assumed.

PROOFS

20. The nature of the various proofs. A number of apparently quite different ways have been developed to carry out the proofs of the formulas for average values of mean squares (see Section 4 for references). They fall quite neatly into two categories:

- (1) Proofs using special machinery or indirect methods (e.g., symmetry arguments and equating of coefficients for special assumptions as in [18] and [20]).

- (2) Proofs using relatively straightforward algebra (e.g., as in [17], [16], and [21]).

While we feel that the first sort of proof offers the real hope for dealing with the more difficult problems which lie ahead, we recognize the usefulness of having examples of the straightforward proofs on record, and have endeavored to keep our proofs direct and to hold the use of special techniques to a minimum.

In the two-way pigeonhole model three samplings take place independently, a sampling of rows, a sampling of columns, and a sampling within cells. The separateness and independence of these samplings is very important in reaching a moderately simple direct proof, as is the possibility of considering the separate samplings as occurring in any order, and then using quantities defined by some array intermediate between the underlying array and the observed array. It is only by combining such a choice with a well-chosen notation and order of procedure that we can keep the algebra from becoming quite heavy.

The original proof [20], was carried through explicitly for the two-way case without replication in the cell ($n = 1$, $N = \infty$), made explicit use of linear models, and depended on two comparisons of three situations:

- (1) rows fixed, columns fixed, within cells sampled,
- (2) rows fixed, columns sampled, within cells sampled,
- (3) rows sampled, columns sampled, within cells sampled.

The next proof [17], found without knowledge of the first result, made explicit use of urn sampling and was carried through explicitly for the two-way case with $\sigma_{ij}^2 = \text{constant}$, and rested most conveniently on the intermediate situation where the within-cell sampling had been completed in each of the RC cells, but the r rows and c columns of the observed array had not yet been fixed. (Note that this requires thinking about, and calculating with, the cell means for the $RC - rc$ cells which will not be observed.) The material in Section 18 is modelled after this proof.

To obtain the average value of the error mean square it is most convenient to think of the sampling as occurring in exactly the reverse order. Here it is most convenient to rest on the intermediate situation where the r rows and c columns have been fixed, but the n individuals to be selected from H in each of these rc cells are still unspecified.

21. The error mean square. The error mean square can be written in various forms, in particular as

$$\frac{1}{rc(n-1)} \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (x_{ijk} - x_{ij.})^2 = \frac{1}{rc} \sum_i \sum_j \left[\frac{1}{n-1} \sum_k (x_{ijk} - x_{ij.})^2 \right].$$

So long as we think of the r rows and c columns as fixed, we have the mean of rc terms of the form

$$\frac{1}{n-1} \sum_k (x_{ijk} - x_{ij.})^2,$$

and the average value of this term is well known from the theory of finite sampling to be $\sigma_{I(i)J(j)}^2$. Thus the average value of the error mean square is

$$\frac{1}{RC} \sum_{i=1}^R \sum_{j=1}^C \sigma_{IJ}^2 = \sigma_R^2.$$

Even though this is clear, we shall give a formal proof as an introduction to the use of a significant technique.

22. Indicator variables. The direct evaluation of the average values of mean squares is greatly facilitated by the use of a simple device [4]. We introduce a set of indicator variables a_1, a_2, \dots, a_R , one for each row in the underlying array. The value of these variables depends on the particular sample of rows which has been selected for the actual array, and is given by:

$$a_I = \begin{cases} 1, & \text{if the } I\text{th row is in the sample of rows,} \\ 0, & \text{otherwise.} \end{cases}$$

From our assumptions about the sampling,

$$\begin{aligned} \text{ave } \{a_I\} &= \text{ave } \{a_I^2\} = \frac{r}{R}, \\ \text{ave } \{a_I a_J\} &= \frac{r}{R} \cdot \frac{r-1}{R-1} = \frac{r(r-1)}{R(R-1)}, \end{aligned}$$

where we write "ave" for the average over all possible samples (we could have written "E" for expectation, but we preferred the more perspicuous notation).

Similarly, we introduce indicator variables b_1, b_2, \dots, b_C for columns, by

$$b_J = \begin{cases} 1, & \text{if the } J\text{th column is in the sample of columns,} \\ 0, & \text{otherwise.} \end{cases}$$

As an illustration of the use of these variables, consider the averaging over all samplings of rows and columns of

$$\frac{1}{rc} \sum_i \sum_j \sigma_{I(i)J(j)}^2 = \frac{1}{rc} \sum_i \sum_j a_i b_j \sigma_{IJ}^2$$

with which we closed the last section. The average value is to be found by replacing a_i by r/R and b_j by c/C , and we have the result announced above.

23. Fundamentals of bisampling. We now consider (i) an arbitrary population of R -by- C arrays $\{y_{IJ}; I = 1, 2, \dots, R; J = 1, 2, \dots, C\}$; (ii) the operation of randomly sampling r rows and c columns; (iii) the resulting r -by- c arrays; and (iv) the row, column, and grand means for these r -by- c arrays. It is our purpose to calculate averages of certain symmetric quadratic expressions in the y_{IJ} , both certain symmetric combinations of variances and covariances of these elements and means, and the averages of certain differences of squares. In the next section knowledge of these combinations will immediately lead us to the formulas for average values of mean squares in a (replicated) two-way classification.

We write, and thus change our convention about capital letters,

$$\begin{aligned} Y_{IJ} &= \text{ave } \{y_{IJ}\}, \\ Y_I &= \frac{1}{C} \sum_J Y_{IJ}, \\ y_I &= \frac{1}{C} \sum_J b_J y_{IJ}, \\ Z_J &= \frac{1}{R} \sum_I Y_{IJ}, \\ z_J &= \frac{1}{r} \sum_I a_I y_{IJ}, \\ Y &= \frac{1}{C} \sum_J Z_J = \frac{1}{R} \sum_I Y_I, \\ y &= \frac{1}{C} \sum_J b_J z_J = \frac{1}{r} \sum_I a_I y_I, \end{aligned}$$

and note that

$$\text{ave } \{y_I\} = Y_I, \quad \text{ave } \{z_J\} = Z_J, \quad \text{ave } \{y\} = Y.$$

We shall find these notations convenient, although we could have written Y_{I-} for Y_I , Y_{-J} for Z_J and Y_{--} for Y . If we had done this, there would have been no convenient parallel notation for the quantities denoted by small letters, since y_I is defined, and will be used, whether or not row I appears in a particular sample of rows, and hence in a particular r -by- c array.

We shall want our results expressed simply. They will have to involve both the average values, Y_{IJ} , and the variances and covariances, of the y_{IJ} . By intuition, or by working out the answer, we can see that they will involve only a few rather symmetric combinations of these moments, namely the three variance components corresponding to the Y_{IJ} :

$$\begin{aligned} \sigma_{\text{row}}^2 &= \frac{1}{R-1} [\sum_I Y_I^2 - RY^2] = \frac{1}{R-1} \sum_I (Y_{I-} - Y_{--})^2, \\ \sigma_{\text{col}}^2 &= \frac{1}{C-1} [\sum_J Z_J^2 - CY^2] = \frac{1}{C-1} \sum_J (Y_{-J} - Y_{--})^2, \\ \sigma_{\text{int}}^2 &= \frac{1}{(R-1)(C-1)} [\sum_I \sum_J Y_{IJ}^2 - C \sum_I Y_I^2 - R \sum_J Z_J^2 + CRY^2] \\ &= \frac{1}{(R-1)(C-1)} \sum_I \sum_J (Y_{IJ} - Y_{I-} - Y_{-J} + Y_{--})^2, \end{aligned}$$

and certain mean variances and covariances, namely,

$$\rho_1 = \frac{1}{RC} \sum_I \sum_J \text{var } \{y_{IJ}\},$$

$$\rho_2 = \frac{1}{RC(C-1)} \sum_I \sum_{j \neq L} \sum \text{cov} \{y_{Ij}, y_{IL}\},$$

$$\rho_3 = \frac{1}{R(R-1)C} \sum_{I \neq K} \sum_j \text{cov} \{y_{Ij}, y_{Kj}\},$$

$$\rho_4 = \frac{1}{R(R-1)C(C-1)} \sum_{I \neq K} \sum_j \sum_{j \neq L} \text{cov} \{y_{Ij}, y_{KL}\}.$$

We begin by writing $y_I y_K$ out as

$$y_I y_K = \frac{1}{c^2} \left[\sum_j b_j^2 y_{Ij} y_{Kj} + \sum_{j \neq L} \sum b_j b_L y_{Ij} y_{KL} \right]$$

and using the independence of the b 's and the y 's in

$$\text{ave} \{b_j b_L y_{Ij} y_{KL}\} = \frac{c(c-1)}{C(C-1)} [Y_{Ij} Y_{KL} + \text{cov} \{y_{Ij}, y_{KL}\}]$$

to find

$$\begin{aligned} \text{ave} \{y_I y_K\} &= \frac{1}{cC} \left[\sum_j Y_{Ij} Y_{Kj} + \frac{c-1}{C-1} \sum_{j \neq L} Y_{Ij} Y_{KL} \right] \\ &\quad + \frac{1}{cC} \left[\sum_j \text{cov} \{y_{Ij}, y_{Kj}\} + \frac{c-1}{C-1} \sum_{j \neq L} \text{cov} \{y_{Ij}, y_{KL}\} \right]. \end{aligned}$$

Now if we use

$$Y_I Y_K = \frac{1}{C^2} \sum_j Y_{Ij} Y_{Kj} + \frac{1}{C^2} \sum_{j \neq L} \sum Y_{Ij} Y_{KL},$$

and $\text{ave} \{y_I\} = Y_I$, $\text{ave} \{y_K\} = Y_K$, and reduce, we have

$$\begin{aligned} \text{cov} \{y_I, y_K\} &= \left(\frac{1}{c} - \frac{1}{C} \right) \frac{1}{C-1} \left[\sum_j Y_{Ij} Y_{Kj} - C Y_I Y_K \right] \\ (1) \quad &\quad + \frac{1}{cC} \left[\sum_j \text{cov} \{y_{Ij}, y_{Kj}\} + \frac{c-1}{C-1} \sum_{j \neq L} \text{cov} \{y_{Ij}, y_{KL}\} \right]. \end{aligned}$$

This is the key result.

If we sum this over I and K with $I \neq K$ and reduce, using

$$\sum_{I \neq K} \sum Y_{Ij} Y_{Kj} + \sum_I Y_{Ij}^2 = R^2 Z_j^2,$$

we find

$$\begin{aligned} &\sum_{I \neq K} \sum \text{cov} \{y_I, y_K\} \\ &= - \left(\frac{1}{c} - \frac{1}{C} \right) \frac{1}{C-1} \left[\sum \sum Y_{Ij}^2 - C \sum Y_I^2 - R^2 \sum Z_j^2 + CR^2 Y^2 \right] \end{aligned}$$

$$\begin{aligned}
 (2) \quad & + \frac{1}{cC} \left[\sum_{I \neq K} \sum_J \text{cov} \{y_{IJ}, y_{KJ}\} + \frac{c-1}{C-1} \sum_{I \neq K} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{KL}\} \right] \\
 & = -\left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 - R(R-1)\sigma_{\text{eol}}^2] \\
 & \quad + R(R-1) \frac{\rho_3 + (c-1)\rho_4}{c}.
 \end{aligned}$$

If we go back to (1), and put $K = I$, we have

$$\begin{aligned}
 (3) \quad \text{var} \{y_I\} & = \left(\frac{1}{c} - \frac{1}{C}\right) \frac{1}{C-1} [\sum_J Y_{IJ}^2 - CY^2] \\
 & \quad + \frac{1}{cC} \left[\sum_J \text{var} \{y_{IJ}\} + \frac{c-1}{C-1} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{IL}\} \right],
 \end{aligned}$$

and summing over I , we find

$$\begin{aligned}
 (4) \quad \sum_I \text{var} \{y_I\} & = \left(\frac{1}{c} - \frac{1}{C}\right) \frac{1}{C-1} [\sum_I \sum_J Y_{IJ}^2 - C \sum_J Y_J^2] \\
 & \quad + \frac{1}{cC} \left[\sum_I \sum_J \text{var} \{y_{IJ}\} + \frac{c-1}{C-1} \sum_I \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{IL}\} \right] \\
 & = \left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 + R\sigma_{\text{eol}}^2] + R \frac{\rho_1 + (c-1)\rho_2}{c} \\
 & = R \left[\left(\frac{1}{c} - \frac{1}{C}\right) \left(1 - \frac{1}{R}\right) \sigma_{\text{int}}^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{eol}}^2 + \frac{1}{c} (\rho_1 + (c-1)\rho_2) \right].
 \end{aligned}$$

When we recall that the relation of y to y_I and Y_I is entirely analogous to that of y_I to y_{IJ} and Y_{IJ} , we see that we can paraphrase (3) to give

$$\begin{aligned}
 (5) \quad \text{var} \{y\} & = \left(\frac{1}{r} - \frac{1}{R}\right) \frac{1}{R-1} [\sum_I Y_I^2 - RY^2] \\
 & \quad + \frac{1}{rR} \left[\sum_I \text{var} \{y_I\} + \frac{r-1}{R-1} \sum_{I \neq K} \text{cov} \{y_I, y_K\} \right] \\
 & = \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{\text{row}}^2 + \frac{1}{rR} \left[\left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 + R\sigma_{\text{eol}}^2] \right. \\
 & \quad + R \frac{\rho_1 + (c-1)\rho_2}{c} \\
 & \quad - \frac{r-1}{R-1} \left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 - R(R-1)\sigma_{\text{eol}}^2] \\
 & \quad \left. + \frac{r-1}{R-1} R(R-1) \frac{\rho_3 + (c-1)\rho_4}{c} \right]
 \end{aligned}$$

$$= \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{\text{row}}^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{col}}^2 + \left(\frac{1}{r} - \frac{1}{R}\right) \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{int}}^2 \\ + \frac{1}{rc} [\rho_1 + (c-1)\rho_2 + (r-1)\rho_3 + (r-1)(c-1)\rho_4].$$

We shall also be interested in the combination,

$$(6) \quad \frac{1}{R} \sum_I \text{var } \{y_I\} - \text{var } \{y\} = -\left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{\text{row}}^2 + \frac{r-1}{r} \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{int}}^2 \\ + \frac{r-1}{rc} [\rho_1 - \rho_3 + (c-1)(\rho_2 - \rho_4)],$$

which becomes, on adding

$$\frac{1}{R} \sum_I (\text{ave } \{y_I\})^2 - (\text{ave } \{y\})^2 = \frac{1}{R} [\sum Y_I^2 - RY^2] = \frac{R-1}{R} \sigma_{\text{row}}^2$$

on corresponding sides,

$$(7) \quad \text{ave } \left\{ \frac{1}{R} \sum y_I^2 - y^2 \right\} \\ = \left(1 - \frac{1}{r}\right) \left[\sigma_{\text{row}}^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{int}}^2 + \frac{1}{c} [\rho_1 - \rho_3 + (c-1)(\rho_2 - \rho_4)] \right].$$

When we observe that the expression for $\sum_J \text{var } \{Z_J\}$ follows from that for $\sum_I \text{var } \{Y_I\}$ by symmetry (interchanging c with r , C with R , and "rows" with "col"), we see that we can easily evaluate another combination

$$(8) \quad \sum_I \sum_J \text{var } \{y_{IJ}\} - C \sum_J \text{var } \{y_J\} - R \sum_I \text{var } \{z_I\} + RC \text{var } \{y\} \\ = RC \left[\left[\left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) - \left(1 - \frac{1}{R}\right) \left(1 - \frac{1}{C}\right) \right] \sigma_{\text{int}}^2 \right. \\ \left. + \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) [\rho_1 - \rho_3 - \rho_2 + \rho_4] \right],$$

which becomes, on adding

$$\sum_I \sum_J (\text{ave } \{y_{IJ}\})^2 - C \sum_J (\text{ave } \{y_J\})^2 - R \sum_I (\text{ave } \{z_I\})^2 + RC (\text{ave } \{y\})^2 \\ = \sum_I \sum_J Y_{IJ}^2 - C \sum_J Y_J^2 - R \sum_I Z_I^2 + RC Y^2 = (R-1)(C-1) \sigma_{\text{int}}^2$$

to corresponding sides,

$$(9) \quad \text{ave } \left\{ \sum_I \sum_J y_{IJ}^2 - C \sum_J y_J^2 - R \sum_I z_I^2 + RC y^2 \right\} \\ = RC \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) [\sigma_{\text{int}}^2 + \rho_1 - \rho_2 - \rho_3 + \rho_4].$$

These formulas will provide us with the desired results. For our limited purposes, they are the fundamentals of bisampling.

24. The replicated two-way classification. In order to develop the formulas for the remaining average mean squares, we shall find it convenient to think of the sampling of n from N as taking place first, and taking place within each of the RC cells of the underlying array, and of the sampling of r rows from R and c columns from C as taking place later. Taking this attitude, we can work with the means determined for each of the RC underlying cells after the sampling within cells and before the sampling of rows and columns. We take Y_{IJ} as the mean for the IJ th cell. We then have

$$\text{ave } \{y_{IJ}\} = Y_{IJ} = X_{IJ-},$$

$$\text{var } \{y_{IJ}\} = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_{IJ}^2,$$

$$\text{cov } \{y_{IJ}, y_{KL}\} = 0, \text{ if } (I, J) \neq (K, L).$$

and

$$\sigma_{\text{row}}^2 = \frac{1}{R-1} \sum_i (X_{i--} - X_{---})^2 = \sigma_R^2,$$

$$\sigma_{\text{col}}^2 = \frac{1}{C-1} \sum_j (X_{-j-} - X_{---})^2 = \sigma_C^2,$$

$$\sigma_{\text{int}}^2 = \frac{1}{(R-1)(C-1)} \sum_i \sum_j (X_{ij-} - X_{i--} - X_{-j-} + X_{---})^2 = \sigma_I^2,$$

$$\rho_1 = \frac{1}{RC} \left(\frac{1}{n} - \frac{1}{N}\right) \sum_i \sum_j \sigma_{IJ}^2 = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_R^2$$

$$\rho_2 = \rho_3 = \rho_4 = 0.$$

The row mean square can be written in various forms, including

$$\begin{aligned} \frac{nc}{r-1} \sum_{i=1}^r (x_{i--} - x_{---})^2 &= \frac{nc}{r-1} \sum_i x_{i--}^2 - \frac{ncr}{r-1} x_{---}^2 \\ &= \frac{nc}{r-1} [\sum_i a_i y_i^2 - r y^2]. \end{aligned}$$

If we use first the independence of a_i from y_i and then (7) of the last section, the average value of this last form becomes

$$\begin{aligned} \frac{nc}{r-1} \text{ave} \left\{ \frac{r}{R} \sum_i y_i^2 - r y^2 \right\} &= \frac{ncr}{r-1} \left(1 - \frac{1}{r}\right) \left[\sigma_R^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_I^2 \right. \\ &\quad \left. + \frac{1}{c} \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_R^2 \right] = nc \sigma_R^2 + n \left(1 - \frac{c}{C}\right) \sigma_I^2 + \left(1 - \frac{n}{N}\right) \sigma_R^2 \end{aligned}$$

as we wished to show. The average value of the columns mean square follows by symmetry.

The interaction mean square can also be written in many forms, some of which are

$$\begin{aligned} & \frac{n}{(r-1)(c-1)} \sum_i \sum_j (x_{ij} - x_{i..} - x_{.j.} + x_{...})^2 \\ &= \frac{n}{(r-1)(c-1)} [\sum_i \sum_j x_{ij}^2 - c \sum_i x_{i..}^2 - r \sum_j x_{.j.}^2 + rcx_{...}^2] \\ &= \frac{n}{(r-1)(c-1)} [\sum_i \sum_j a_i b_j y_{ij}^2 - c \sum_i a_i y_i^2 - r \sum_j b_j z_j^2 + rcy^2]. \end{aligned}$$

If again we use the independence of a_i and b_j from each other and the other quantities, and then (9) of the last section, the average value of the last form is seen to be

$$\begin{aligned} & \frac{n}{(r-1)(c-1)} \left[\frac{rc}{RC} \sum_i \sum_j Y_{ij}^2 - \frac{rc}{R} \sum_i Y_i^2 - \frac{rc}{C} \sum_j Z_j^2 + rcY^2 \right] \\ &= \frac{n}{(r-1)(c-1)} RC \frac{rc}{RC} \left(1 - \frac{1}{r} \right) \left(1 - \frac{1}{c} \right) \left[\sigma_i^2 + \left(\frac{1}{n} - \frac{1}{N} \right) \sigma_{\#}^2 \right] \\ &= n\sigma_i^2 + \left(1 - \frac{n}{N} \right) \sigma_{\#}^2 \end{aligned}$$

as we wished to show. Thus we have the average values of the mean squares for the two-way pigeonhole model with replication.

25. The unreplicated three-way classification. To deal with the three-way pigeonhole model, where r rows from R , c columns from C , and s slices from S are independently sampled, we start out to calculate the average mean squares for rows, columns, and their interaction just as for the replicated two-way. Differences will first appear when we come to calculate ρ_1 , ρ_2 , ρ_3 , and ρ_4 , which is most simply done in an indirect way.

We remark that, now, we have

$$\begin{aligned} y_{IJ} &= \frac{1}{s} \sum_{k=1}^s x_{IJK(k)} = \frac{1}{S} \sum_{K=1}^S c_K x_{IJK}, \\ \text{ave } \{y_{IJ}\} &= Y_{IJ} = \frac{1}{S} \sum_K x_{IJK}, \end{aligned}$$

where $\{c_K\}$ are a new set of indicator variables which specify the sampling of s slices from S . Now

$$\text{var } \{y_{IJ}\} = \left(\frac{1}{s} - \frac{1}{S} \right) \frac{1}{S-1} [\sum_K x_{IJK}^2 - Sx_{IJ.}^2]$$

since we have a sample mean of s from S , so that

$$RC\rho_1 = \sum_I \sum_J \text{var} \{y_{IJ}\} = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} [\sum_I \sum_J \sum_K x_{IJK}^2 - S \sum_I \sum_J x_{IJ-}^2].$$

Moreover

$$\begin{aligned} \text{var} \left\{ \sum_J y_{IJ} \right\} &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} [\sum_K (\sum_J x_{IJK})^2 - S(\sum_J x_{IJ-})^2] \\ &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} [C^2 \sum_K x_{I-K}^2 - C^2 S x_{I--}^2]. \end{aligned}$$

Since $\sum_J y_{IJ}$ is the mean of a sample of s from the S values $\sum_J x_{IJK}$, and hence

$$\begin{aligned} RC\rho_1 + RC(C-1)\rho_2 &= \sum_I \text{var} \left\{ \sum_J y_{IJ} \right\} \\ &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{C^2}{S-1} [\sum_I \sum_K x_{I-K}^2 - S \sum_I x_{I--}^2]. \end{aligned}$$

By symmetry, then

$$RC\rho_1 + R(R-1)\rho_3 = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{R^2}{S-1} [\sum_J \sum_K x_{-JK}^2 - S \sum_J x_{-J-}^2]$$

and, indeed,

$$\begin{aligned} RC\rho_1 + RC(C-1)\rho_2 + R(R-1)\rho_3 + R(R-1)C(C-1)\rho_4 \\ = \text{var} \left\{ \sum_I \sum_J y_{IJ} \right\} = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} [\sum_K (\sum_I \sum_J x_{IJK})^2 - S(\sum_I \sum_J x_{IJ-})^2] \\ = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{C^2 R^2}{S-1} [\sum_K x_{--K}^2 - S x_{---}^2]. \end{aligned}$$

We now have the basis for evaluating the ρ 's.

We introduce the additional variance components for the three-way classification by

$$\sigma_s^2 = \frac{1}{S-1} [\sum_K x_{--K}^2 - S x_{---}^2],$$

$$\sigma_{RS}^2 = \frac{1}{(R-1)(S-1)} [\sum_I \sum_K x_{I-K}^2 - R \sum_K x_{-K-}^2 - S \sum_I x_{I--}^2 + R S x_{---}^2].$$

$$\sigma_{CS}^2 = \frac{1}{(C-1)(S-1)} [\sum_J \sum_K x_{-JK}^2 - C \sum_J x_{-J-}^2 - S \sum_J x_{-J-}^2 + C S x_{---}^2],$$

and finally,

$$\begin{aligned} \sigma_{RCS}^2 &= \frac{1}{(R-1)(C-1)(S-1)} [\sum_I \sum_J \sum_K x_{IJK}^2 - R \sum_J \sum_K x_{-JK}^2 - C \sum_I \sum_K x_{I-K}^2 \\ &\quad - S \sum_I \sum_J x_{IJ-}^2 + RC \sum_K x_{--K}^2 + RS \sum_J x_{-J-}^2 + CS \sum_I x_{I--}^2 - RCS x_{---}^2]. \end{aligned}$$

In terms of these quantities we can now calculate the values of the ρ 's. We can conveniently use

$$\begin{aligned}\sum_i \sum_K x_{i-K}^2 - S \sum_i x_{i--}^2 &= (R-1)(S-1)\sigma_{RS}^2 + R[\sum_K x_{--K}^2 - Sx_{--}^2] \\ &= (R-1)(S-1)\sigma_{RS}^2 + R(S-1)\sigma_S^2,\end{aligned}$$

and, by symmetry,

$$\sum_j \sum_K x_{-jK}^2 - S \sum_j x_{-j-}^2 = (C-1)(S-1)\sigma_{CS}^2 + C(S-1)\sigma_S^2,$$

and, by a similar though longer calculation,

$$\begin{aligned}\sum_i \sum_j \sum_K x_{ijK}^2 - S \sum_i \sum_j x_{ij-}^2 &= (R-1)(C-1)(S-1)\sigma_{RCS}^2 \\ &+ R(C-1)(S-1)\sigma_{CS}^2 + (R-1)C(S-1)\sigma_{RS}^2 + RC(S-1)\sigma_S^2.\end{aligned}$$

Substituting these into the earlier formulas, and removing factors where convenient, yields

$$\rho_1 = \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{1}{R}\right) \left(1 - \frac{1}{C}\right) \sigma_{RCS}^2 + \left(1 - \frac{1}{C}\right) \sigma_{CS}^2 + \left(1 - \frac{1}{R}\right) \sigma_{RS}^2 + \sigma_S^2 \right],$$

$$\rho_1 + (C-1)\rho_2 = \left(\frac{1}{S} - \frac{1}{S}\right) C \left[\left(1 - \frac{1}{R}\right) \sigma_{RS}^2 + \sigma_S^2 \right],$$

$$\rho_1 + (R-1)\rho_3 = \left(\frac{1}{S} - \frac{1}{S}\right) R \left[\left(1 - \frac{1}{C}\right) \sigma_{CS}^2 + \sigma_S^2 \right],$$

$$\rho_1 + (C-1)\rho_2 + (R-1)\rho_3 + (C-1)(R-1)\rho_4 = \left(\frac{1}{S} - \frac{1}{S}\right) RC\sigma_S^2.$$

From these we could compute all the ρ 's, but it is simpler to find the combinations which will be of most use to us. These are

$$\rho_1 - \rho_3 = \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{1}{C}\right) \sigma_{RCS}^2 + \sigma_{RS}^2 \right],$$

$$\rho_2 - \rho_4 = \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{1}{C}\right) \sigma_{RCS}^2 + \sigma_{RS}^2 \right],$$

whence

$$\rho_1 - \rho_3 + (C-1)(\rho_2 - \rho_4) = \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{C}{C}\right) \sigma_{RCS}^2 + \sigma_{RS}^2 \right],$$

$$\rho_1 - \rho_2 - \rho_3 + \rho_4 = \left(\frac{1}{S} - \frac{1}{S}\right) [\sigma_{RCS}^2].$$

We are now ready to return to the evaluations of Section 23, where we found the average mean squares for rows to be

$$nc\sigma_R^2 + n \left(1 - \frac{C}{C}\right) \sigma_I^2 + n[\rho_1 - \rho_3 + (C-1)(\rho_2 - \rho_4)].$$

We have now to write s in place of n , σ_{RC}^2 in place of σ_I^2 , and to substitute in the value we have just found for quantity in brackets. We obtain, then, for the average value of a row mean square in a pigeonhole three-way

$$cs\sigma_R^2 + s\left(1 - \frac{c}{C}\right)\sigma_{RC}^2 + c\left(1 - \frac{s}{S}\right)\sigma_{RS}^2 + \left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2.$$

The average values of column and slice mean squares now follow by symmetry.

We had obtained

$$n\sigma_I^2 + n[\rho_1 - \rho_2 - \rho_3 - \rho_4]$$

for the average value of the interaction mean square. Making the necessary changes, the average value of the RC -interaction mean square for the three-way pigeonhole becomes

$$s\sigma_{RC}^2 + \left(1 - \frac{s}{S}\right)\sigma_{RCS}^2,$$

and the average values for RS and CS interactions follow by symmetry. There remains the triple interaction.

26. The triple interaction. To get a simple hold on the triple interaction, we will find it convenient to introduce some differencing operators with the following definitions:

$$\delta_{ii'} f(i, j, k) = \frac{1}{\sqrt{2}} (f(i, j, k) - f(i', j, k)),$$

$$\Delta_{II'} f(I, J, K) = \frac{1}{\sqrt{2}} (f(I, J, K) - f(I', J, K)).$$

and similarly for other indices. The usefulness of these operators stems from the following chain of representations for interaction sums of squares:

$$\sum_i x_{i..}^2 - rx_{...}^2 = \frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_{i..})^2,$$

$$\sum_i \sum_j x_{ij.}^2 - r \sum_j x_{.j.}^2 - c \sum_i x_{i..}^2 + rcx_{...}^2 = \frac{1}{rc} \sum_i \sum_{i'} \sum_j \sum_{j'} (\delta_{ii'} \delta_{jj'} x_{ij.})^2.$$

$$\sum_i \sum_j \sum_k x_{ijk}^2 - r \sum_j \sum_k x_{.jk}^2 - c \sum_i \sum_k x_{i..k}^2 - s \sum_i \sum_j x_{ij.}^2 + rc \sum_k x_{..k}^2$$

$$+ rs \sum_j x_{.j.}^2 - cs \sum_i x_{i..}^2 - rcx_{...}^2 = \frac{1}{rcs} \sum_i \sum_{i'} \sum_j \sum_{j'} \sum_k \sum_{k'} (\delta_{ii'} \delta_{jj'} \delta_{kk'} x_{ijk})^2.$$

and so on.

To establish these representations we have only to prove (as we may by direct expansion), that

$$\sum_{h=1}^m u_h^2 - mu^2 = \frac{1}{m} \sum_h \sum_{h'} (\delta_{hh'} u_h)^2$$

and apply induction, writing, for example, the second interaction in the forms

$$\begin{aligned} \sum_j [\sum_i x_{ij}^2 - rx_{.j}^2] - c[\sum_i x_{i..}^2 - rx_{...}^2] \\ = \sum_j \left[\frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_{ij.})^2 \right] - c \left[\frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_{i..})^2 \right] \\ = \sum_i \sum_{i'} \frac{1}{r} \left[\sum_j (\delta_{ii'} x_{ij.})^2 - c(\delta_{ii'} x_{i..})^2 \right] \\ = \sum_i \sum_{i'} \frac{1}{r} \left[\sum_j \sum_{j'} \frac{1}{c} (\delta_{jj'} \delta_{ii'} x_{ij.})^2 \right], \end{aligned}$$

where we have used the identity twice and the fact that the mean value of $\delta_{ii'} x_{ij.}$ over j is $\delta_{ii'} x_{i..}$ itself.

These representations lead at once to average value formulas. Thus, in the one-way case

$$\begin{aligned} \text{ave} \left\{ \sum_i x_i^2 - rx_{.}^2 \right\} &= \text{ave} \left\{ \frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_i)^2 \right\} = \text{ave} \left\{ \frac{1}{r} \sum_i \sum_{i'} a_i a_{i'} (\Delta_{ii'} x_i)^2 \right\} \\ &= \frac{1}{r} \sum_i \sum_{i'} \frac{r(r-1)}{R(R-1)} (\Delta_{ii'} x_i)^2 = \left(\frac{r-1}{R-1} \right) \frac{1}{R} \sum_i \sum_{i'} (\Delta_{ii'} x_i)^2 \\ &= \frac{r-1}{R-1} [\sum_i x_i^2 - Rx_{.}^2], \end{aligned}$$

where we have used the fact that $a_i a_{i'}$ equals $r(r-1)/R(R-1)$ except when $I = I'$, and the fact that we can neglect $I = I'$ because $\delta_{II'} x_i$ vanishes for $I = I'$. Similarly, in a two-way case

$$\begin{aligned} \text{ave} \left\{ \sum_i \sum_j x_{ij}^2 - r \sum_j x_{.j}^2 - c \sum_i x_{i..}^2 + rcx_{...}^2 \right\} \\ = \text{ave} \left\{ \frac{1}{rc} \sum_i \sum_{i'} \sum_j \sum_{j'} (\delta_{ii'} \delta_{jj'} x_{ij.})^2 \right\} \\ = \frac{1}{rc} \text{ave} \left\{ \sum_i \sum_{i'} \sum_j \sum_{j'} a_i a_{i'} b_j b_{j'} (\Delta_{ii'} \Delta_{jj'} x_{ij.})^2 \right\} \\ = \frac{1}{rc} \frac{r(r-1)}{R(R-1)} \frac{c(c-1)}{C(C-1)} \sum_i \sum_{i'} \sum_j \sum_{j'} (\Delta_{ii'} \Delta_{jj'} x_{ij.})^2 \\ = \frac{(r-1)(c-1)}{(R-1)(C-1)} [\sum_i \sum_j x_{ij}^2 - R \sum_j x_{.j}^2 - C \sum_i x_{i..}^2 + RCx_{...}^2] \end{aligned}$$

and so on. The next case, for the three-way case, shows us that the average value of the RCS -interaction is exactly σ_{RCS}^2 as we wished to prove.

Clearly the extensions of this argument to the four-way, five-way, etc., classifications will always give a similar answer for the lowest interaction (the one involving all the indices!) for any factorial.

27. Further generalization. It is just a matter of moderately stiff algebra to carry on an inductive proof of similar results for more classifications. If $\text{ave } \{y_{IJK}\} = Y_{IJK}$, and if $q(\{y_{ijk}\})$ is a quadratic in the observed y_{ijk} , then $\text{ave } \{q(\{y_{ijk}\})\} = \text{ave } \{q(\{Y_{ijk}\})\}$

+ (a linear form in variances and covariances of the $\{y_{ijk}\}$)

Now the capital Y 's correspond to an analysis without replication, so that, for the $q(\{Y_{ijk}\})$ we are interested in, the averages on the right can be calculated from the results of the last section. Because of symmetry, the linear form in variances and covariances will be a linear combination of 8 quantities, the mean variance, and seven mean covariances. Working these expressions out, and then calculating the values of these mean variance and covariances in special cases, we can obtain the average values of mean squares for the replicated three-way and unreplicated four-way designs.

After this we are ready for another step of the induction, and so on.

Clearly systematic algebra can take us deep into the forest of notation. But the detailed manipulation will, sooner or later, blot out any understanding we may have started with. If there is a way of seeing some aspects of the final result more directly, then it will be worth while to seize it.

There are a number of such ways involving special tools or devices of varying complexity. Since we have tried to keep the approach of this paper reasonably pedestrian (although indicator variables and the $\delta_{ii'}$ and $\Delta_{II'}$ may be regarded as the equivalent of roller skates!), we shall try to use the least special way that we know.

28. More direct insight. Let us ask about the coefficient with which

$$\sigma_{RCDEFGS}^2$$

appears in the average value of the mean square of the $RCDE$ -interaction in an 11-way design with factors labelled $R, C, D, E, F, G, H, J, K, L, S$. More specifically, let us ask how the coefficient depends on the values of s and S . It will appear that we can answer this rather directly.

There will exist some formulas which make up the fundamentals of decim-sampling. Apply them to the 10-way classification involving $R, C, D, E, F, G, H, J, K, L$ —involving all the classifications but S . They will give the average value of the $RCDE$ -interaction in terms of σ^2 's whose indices do not include S and of a mean variance and $2^{10} - 1 = 1023$ mean covariances. The latter 1024 quantities will be expressible in terms of the σ^2 's which involve S by a process entirely similar (though more complicated in detail) than that used in Section 15, each of the 1024 formulas for a linear combination of mean variance and covariances will involve a factor

$$\left(\frac{1}{s} - \frac{1}{S}\right),$$

since, in every case, we shall be sampling s out of S . Hence, when all the algebraic

dust has quieted down, the coefficient of $\sigma_{RCDEFGS}^2$ will include the same factor, and will depend in no other way on S . It will depend on s in a further way, since finding the row, column, D -variable, and E -variable will still leave us with $fghjkl$ different cells. Thus a factor of this value will appear in the mean square for $RCDE$ when that mean square is written out in terms of means (not totals). Thus the complete dependence on s and S will be

$$s \left(\frac{1}{s} - \frac{1}{S} \right) = 1 - \frac{s}{S}.$$

With this result, and the result, obtained incidentally, that h and H will enter only through a factor h , we are essentially finished.

By symmetry, we see that the term in the mean-value of the $RCDE$ -interaction with which we are concerned is

$$(\text{constant}) \left(1 - \frac{f}{F} \right) \left(1 - \frac{g}{G} \right) h j k l \left(1 - \frac{s}{S} \right) \sigma_{RCDEFGS}^2$$

If we simplify matters by choosing $h = H = j = J = k = K = l = L = 1$, the whole 11-way classification condenses to a 7-way classification and if, moreover, we revert to Model II (everything independent and normal), we obtain

$$(\text{constant}) \sigma_{RCDEFGS}^2$$

as a term in the average value of the $RCDE$ -mean square, and it is well known that the constant must now be unity.

This provides a not-too-indirect proof for the rules set forth in Section 11.

REFERENCES

- [1] R. L. ANDERSON AND T. A. BANCROFT, *Statistical Theory in Research*, McGraw-Hill, New York, 1952.
- [2] C. A. BENNETT AND N. L. FRANKLIN, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley, New York, 1954.
- [3] W. G. COCHRAN, "The use of analysis of variance in enumeration by sampling," *J. Amer. Statist. Assoc.*, Vol. 34 (1939), pp. 492-510.
- [4] J. CORNFIELD, "On samples from finite populations," *J. Amer. Statist. Assoc.*, Vol. 3 (1944), pp. 236.
- [5] C. EISENHART, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1-21.
- [6] M. H. HANSEN, W. N. HURWITZ AND W. G. MADOW, *Sample Survey Methods and Theory*, Vol. 2, *Theory*, John Wiley, New York, 1953.
- [7] W. A. HENDRICKS, "Variance components as a tool for the analysis of sample data," *Biometrics*, Vol. 7 (1951), pp. 97-101.
- [8] O. KEMPTHORNE, *Design and Analysis of Experiments*, John Wiley, New York, 1952.
- [9] O. KEMPTHORNE, "The randomization theory of experimental inference," *J. Amer. Statist. Assoc.*, Vol. 50 (1955), pp. 946-967.
- [10] M. D. MCCARTHY, "On the application of the z -test to randomized blocks," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 337-359.
- [11] A. M. MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill, 1950.
- [12] J. NEYMAN WITH THE COOPERATION OF K. IWASKIEWICZ AND ST. KOŁODZIECZYK, "Statistical problems in agricultural experimentation," *J. Roy. Stat. Soc., Suppl.*, Vol. 2 (1935), pp. 107-180.

- [13] E. J. G. PITMAN, "Significance tests which may be applied to samples from any population, III. The analysis of variance test," *Biometrika*, Vol. 29 (1938), pp. 322-335.
- [14] J. W. TUKEY, "Dyadic anova, an analysis of variance for vectors," *Human Biology*, Vol. 21 (1949), pp. 65-110.
- [15] B. L. WELCH, "On the z -test in randomized blocks and Latin squares," *Biometrika*, Vol. 29 (1937), pp. 21-52.
- [16] M. B. WILK, *The Randomization Analysis of a Generalized Randomized Block Design*, M.S. Thesis, Iowa State College 1953.

RESEARCH REPORTS

- [17] J. CORNFELD, "The two-way classification in the analysis of variance from the point of view of urn sampling," mimeo from Office of Biometry, National Institutes of Health (ca. 1953).
- [18] R. HOOKE, "Symmetric functions of a two-way array," *Ann. Math. Stat.* Vol. 27 (1956), pp. 55-79.
- [19] O. KEMPTHORNE, dittoed early draft of book on *Genetic Statistics*, used in class about November, 1953.
- [20] J. W. TUKEY, "Interaction in a row-by-column design," Memorandum Report 18, Statistical Research Group, Princeton University, July, 1949.
- [21] M. B. WILK, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance," Report AV-3 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, December, 1953.
- [22] M. B. WILK, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, II. On the analysis of a general balanced randomized block type design," Report AV-4 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, March, 1954.
- [23] M. B. WILK AND O. KEMPTHORNE, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, III. The expectation of mean square for a general three way classification," Report AV-7 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, June, 1954.
- [24] M. B. WILK AND O. KEMPTHORNE, "Fixed, mixed, and random models," *J. Amer. Statist. Assoc.*, Vol. 50 (1955), pp. 1144-1167. "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, IV. Fixed, mixed, and random models," Report AV-8 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, August, 1954. This report covered by a paper submitted to the *J. Amer. Stat. Assn.*
- [25] M. B. WILK, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, V. The expectation of mean squares for a general three-way classification (Part 2)," Report AV-9 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, August, 1954.
- [26] M. B. WILK AND O. KEMPTHORNE, "Derived linear models and their use in the analysis of randomized experiments," (Final report, covering [21] through [25], Analysis of Variance Project, Statistical Laboratory, Iowa State College, April, 1955.

SOME ASPECTS OF THE ANALYSIS OF FACTORIAL EXPERIMENTS IN A COMPLETELY RANDOMIZED DESIGN¹

BY M. B. WILK² AND O. KEMPTHORNE³

Iowa State College and Princeton University; Iowa State College

1. Introduction. This paper is concerned with some aspects of the statistical analysis of factorial experiments carried out according to a completely randomized design, and is one of the joint portions of an investigation into the role and meaning of linear statistical models in the analysis of randomized experiments.

There are essentially two ways of obtaining the analysis of data obtained in a comparative experiment. One way, which is given in standard texts, is to write down a model of the type

$$y_{ijk\dots} = \mu + a_i + b_j + \dots \text{etc.},$$

where $y_{ijk\dots}$ is the observation and the terms on the right-hand side are fixed unknown constants or random variables with specified properties. The above equation with a complete statement of all the properties of the quantities contained in it is usually called *the* model for the experiment. The texts and the literature are to the best of our knowledge, with a few exceptions to be mentioned later, bare with regard to how one determines the model, how one answers a question such as "Why not a multiplicative model?" or "Why are the a 's fixed and the b 's random?" The other way is that practiced intuitively by many experimental statisticians and described most aptly by Fisher ([3], [4], [5], [6]) in which (a) one envisages an analysis of variance of the observations from the point of view of topography, apart from treatment, such as for instance in a field experiment by rows, columns, plots within row-column cells, etc.; (b) one envisages an analysis of variance by treatments; (c) one notes how the treatments have been assigned to the experimental material, such as, for instance, factor α to rows; and (d) one therefore sees with which part of the topographical analysis any particular component of the treatment breakdown should be associated.

The second procedure cannot be regarded as fully specified by what is said above. The first procedure can only be regarded as arbitrary unless some logical basis can be given for it. It is to the problem implied in the last sentence which we have addressed our work.

In preparing this paper for publication we have had the benefit of specific

Received December 30, 1955.

¹ This paper is based on a part of the Ph.D. Thesis (Iowa State College, 1955) of the senior author [14].

² Some of the investigations on which this paper is based were sponsored by Wright Air Development Center. Part of the preparation of this paper for publication was in connection with research sponsored by the Office of Naval Research.

³ Journal Paper No. 3057 of the Iowa State College Agricultural Experiment Station.

and general criticism and suggestions from Professor John Tukey, whose assistance and advice it is a pleasure to acknowledge.

2. Relation to other work. The general history of the line of attack is given by Wilk and Kempthorne [15]. Since that time Smith [10], Scheffe [8], and Cornfield and Tukey [1] have worked on the general problems indicated above. Cornfield and Tukey [1] also discuss relations between approaches to the problem.

3. Some fundamental concepts. The concepts implied in the words "treatment" (or "factor level"),⁴ "experimental unit" and "true response" enter importantly into the developments in the sequel. We shall attempt to convey the general meaning which these terms have for us.

While recognizing that the term treatment generally (operationally) designates a *category* of entities or operations, we shall use it as synonymous with "ideal treatment" or "typical treatment." An example of treatment as a category is a variety of, say, corn with operational representations as individual seeds, so that the treatment may be thought of as having a nested structure. The conception of a treatment such as "a temperature of 45°C" is often different. Even if temperature control is difficult, so that in an actual trial one uses $(45^\circ + \epsilon)^\circ\text{C}$ with ϵ unknown, one usually feels that it is reasonable to conceive, at least on a macroscopic scale, of a "true or ideal treatment" of 45°C, in the attainment of which we are frustrated by physical difficulties.

In most cases it is useful to introduce explicitly the notion of a "treatment error" which will reflect the difficulty in attaining or reproducing a conceptually meaningful ideal. In this paper we shall take such a view. The case when the treatment should properly be regarded as a member of a well-defined population will be given in a later paper.

A reasonable operational definition of experimental units, though circular to some extent, is "those entities in an experiment to which treatments are assigned at random." It is often possible and useful to think of experimental units as physical entities such as plots of land or individual animals, but in many cases such a view is misleadingly naive. Extensions of the term to include periods of time, states of mind, and other ill-defined complexes of conditions are needed. In an agronomic experiment we would regard the unit not simply as a plot of ground, but rather as the plot plus weather and other conditions not subject to test. In specific instances such a view involving "ultimate identification" of experimental units may be too restrictive and could be meaningfully and usefully relaxed. In the formal developments in the sequel, we shall be operationally deterministic in that we shall regard an experimental unit to be conceptually entirely identified so that a given stimulus would produce a definite response. This should not be construed to mean that every situation must be fitted exactly into such a context for the analysis to be useful.

⁴ In general a "treatment" is partially specified by a "factor level." However, most of our remarks can be read substituting "factor level" for "treatment."

The notion of "true" or "typical" response seems readily meaningful at least superficially, and deeper analysis immediately involves one in philosophical discussion which is unnecessary in most experimental contexts.

As regards "experimental error" it may be useful to distinguish between "physical errors" and "sampling errors"; and in the first category to distinguish the experimenter's concern with "systematic errors" from the statistician's treatment which usually revolves around an assumption of "random errors." Some obvious categories of physical errors, with respect to subjecting a given experimental unit to a given treatment and observing the response, are errors of measurement of the response, errors of treatment application, and, from some points of view, errors dependent on the "physical state" of the experimental unit. As we shall see, certain sampling errors can be controlled, in a statistical sense, by the device of randomization. In the analysis of other errors the statistician and experimenter must rely on judicious assumption based on insight and experience.

4. The experimental situation and design; basic notation. The essence of the completely randomized design is that no attempt is made to structure the experimental units; or from another, more accurate viewpoint, no restrictions are imposed in the random assignment of treatments to available experimental units.

We shall describe in detail a situation in which treatment combinations of interest may be classified according to the "levels" of three "factors." This will provide enough generality to indicate extension of the methods and results. The case of two factors can be obtained formally by considering one factor to have only one level.

The factors (e.g., temperature, varieties, types of acid, etc.) will be denoted by script letters \mathcal{A} , \mathcal{B} , \mathcal{C} . The number of levels of each factor, in the experimental population, will be denoted by the corresponding capital letters A , B , C . We suppose, for purpose of reference only, that the levels of each factor are ordered (arbitrarily) and let $i = 1, 2, \dots, A$; $j = 1, 2, \dots, B$; $k = 1, 2, \dots, C$, denote the various levels in the populations of levels of factors \mathcal{A} , \mathcal{B} and \mathcal{C} , respectively.

Suppose there are P experimental units with respect to which we wish to study comparatively the various treatment combinations. (The details of *what* we may be interested in doing will vary with the specific physical situation, but some general statistical aspects of what the bare situation and design enable us to do remain the same.) Again we suppose, for formal reference, that the units are ordered, and let $m = 1, 2, \dots, P$ denote the unit in the population of units.

The experimental design is now defined as follows:

- (i) Select a levels from A of factor \mathcal{A} at random.
- (ii) Select b levels from B of factor \mathcal{B} at random.
- (iii) Select c levels from C of factor \mathcal{C} at random.

(We will use the notation $i^* = 1, 2, \dots, a$; $j^* = 1, 2, \dots, b$; $k^* = 1, 2, \dots, c$ to denote the randomly selected levels of \mathcal{A} , \mathcal{B} , and \mathcal{C} respectively *in order of*

their selection. Thus, for example, $i^* = 1$ corresponds with probability $1/A$ to any designated value of i .)

(iv) Select p experimental units at random from P , where

$$p = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c n_{i \cdot j \cdot k}, \quad \text{and all } n_{i \cdot j \cdot k} \geq 1.$$

The values of the $n_{i \cdot j \cdot k}$ are treated as known prechosen fixed numbers. (Further explanation of this is given at the end of this section.)

(v) Apply selected factor levels to selected experimental units at random but so that selected treatment combination ($i^*j^*k^*$) appears on $n_{i \cdot j \cdot k}$ of the selected experimental units.

Some purely formal difficulties can arise in this general exposition for the case of, say, $a = A$. According to our description above, the identification of levels of α by i^* would be a random arrangement of that effected by i . In dealing with symmetric functions, clearly no difficulties arise. The whole matter can be handled simply by a convention that, for example, when $A = a$ we take i and i^* to be identical indices; or, where non-symmetric functions arise, it can be handled by an extended notation, as will be seen in the sequel.

It is most natural to think of the design as being imposed upon given background populations of levels of factors and of experimental units, but it should be pointed out that it is in fact our procedure in the design which determines the relevant (statistically) population of treatments and units to which our experiment applies. Some further discussion of consequences of this point is given below.

The description above is intended to be general. Cases of fixed, mixed, and random model situations are included as special cases. The possibility of equal, proportional, or unequal numbers in the "subclasses" of the observations is allowed for. In the described set-up the number of observations associated with a treatment combination depends on the actual realization of the experiment, that is, on the outcome of the random selections, and not, in general, on the population of treatments. An important exception to this is the case of fixed factors. Thus, we specify that the *selected* treatment combination ($i^*j^*k^*$) appear $n_{i \cdot j \cdot k}$ times; but, in general, the association of ($i^*j^*k^*$) with values of (ijk) will depend on the random selection process.

5. Some discussion. In the formal description of the experimental situation and design in the preceding section, the role of experimental units in the experimental situation and the relation of the sample of units to the population are specified explicitly.

The population of units defines, in a sense, our experimental milieu or background. Even if all units can be thought of as identical (a rare event) many background influences (not under direct study) are being "held constant." For example, 10 cc samples from a well-mixed, non-volatile solution may well be considered (aside from pipetting errors) essentially identical. But if, in a two-factor experiment, one factor consists of levels of concentration of a reagent and another

is time allowed for reaction, then it is part of the relevant background for inference, tied up with the conception of experimental unit, to describe (or at least keep in mind the existence of) such influences as ambient temperature, barometric pressure, type and shape of container, etc. Thus our inferences must always be interpreted with respect to some population of experimental units, even though in specific instances we may be quite certain of the absence of influence of certain aspects of the experimental milieu.

The emphasis on the relation of sample to population is a fundamental contribution of procedures of modern statistical inference toward scientific objectivity. In spite of the wide acceptance which, we believe, the preceding sentence would find there appears to be some tendency among statisticians to think of the population to which *statistical* inferences are to be made to be not that from which the random sample is obtained but rather one which is indicated by their "interest." The key to this difficulty may lie in the failure to recognize any distinctions between "empirical inferences" based on statistical techniques and "scientific inferences" based on theories of mechanism, mechanical analogies, intuition, etc., in *addition* to statistical inferences.

For example, in a two-factor experiment involving specific insecticides tested with respect to a random selection of 15 types of insects from a population of 200 types of insects we would recognize the statistical validity of two viewpoints in evaluating the comparative utility of the insecticides: (i) relative to the entire population of 200 types of insects from which we have a random sample; (ii) relative to the 15 types of insects actually tested (i.e., the ones which appeared in our random selection). There does not appear to be any *general* justification in attempting, on the basis of data relating to 15 non-randomly selected types of insects (as, for example, those prominent in a certain region), to extend the statistical (empirical) inference to some broader, undefined, population of insect types. There can be no question as to the need or importance of making such an extension, but such extension is essentially non-statistical and must be based on subject matter knowledge and intuition.

6. A conceptual framework for analyses; the population model. In the previous sections we have described an experimental situation and procedure which, at least formally (and granted agreement on the meaning and necessary procedure implied by "random"), is non-controversial. We propose now to provide a conceptual framework for the statistical analysis. This will naturally require some assumptions, all of which we will attempt to make elementary, in the sense that their implications are easily appreciated, and explicit.

We postulate the existence of a real (unknown number Y_{ijkm} which represents the "true" (or "typical") response if unit m is subjected to the treatment combination consisting of the i th level of \mathcal{A} , j th level of \mathcal{B} , and k th level of \mathcal{C} ; and we take as our immediate framework of statistical concern the conceptual set $\{Y_{ijkm}\}$, and more particularly certain functions defined on the elements of this set. Several presumptions are implicit in the preceding sentences. First, the scale of observation is considered as "given," though our succeeding discussion could

proceed equally well in terms of any function of Y . This is *not* to imply that any scale for analysis is as informative as any other. Second, for the quantity Y_{ijkm} to have meaning by itself it is a necessary assumption that the response from given treatment on given unit be dependent only on that treatment and unit alone, and not on the over-all configuration of other treatments and other units; this excludes certain experiments such as those involving competition in animal-feeding trials. Third, we assume that the notion of "true" or "typical" response can be given an objective meaning in the given situation.

Proceeding now on the basis of the previous paragraph, we know that if we actually subjected unit m to factor combination (ijk) , we would not in general observe the true response, Y_{ijkm} , owing to inevitable errors in treatment application, in response measurement, and variations for a given unit owing to its "physical state". These types of errors we refer to as technical errors. These technical errors have no relation to the formal randomization procedure but belong to the conceptual framework. Consequently, in a general study of this sort we have three alternatives with respect to technical errors: (i) To deal with the "ideal" case where such errors are not considered, with the understanding that the application of the method and results in specific situations would require some extensions, depending on "reasonable" assumptions in the specific case. (ii) To employ simple assumptions, which are popular, easily understood, and often reasonable, again with the understanding that adjustment may be necessary to meet specific situations. (iii) To attempt to carry technical errors with some sort of "maximum generality." Procedure (ii) appeared to us to be the most useful.

Accordingly, we will assume that if combination (ijk) were applied to unit m , then we would observe

$$y_{ijkm} = Y_{ijkm} + \epsilon_{ijkm},$$

where the ϵ_{ijkm} , representative of combined technical errors, can be treated as random variables which are mutually uncorrelated with mean 0 and common variance σ^2 .

Some directions of increasing generality of assumptions would be (i) relaxing the homogeneity assumptions to, say, variance $(\epsilon_{ijkm}) = \sigma_m^2$; (ii) relaxing the homogeneity assumptions to, say, variance $(\epsilon_{ijkm}) = \sigma_i^2$; (iii) y_{ijkm} follows some distribution $F_{ijkm}(y)$ of which Y_{ijkm} is some parameter. It is easy to see that the results we shall give are in fact essentially valid if generalization (i) above is permitted; we have not built it in explicitly to simplify the presentation and lay clear some aspects of the results. Furthermore, the results on ems (expectation of mean squares) are essentially valid if generalization (ii) is allowed.

Anticipating its utility in the succeeding section we can now write down the population model as

$$y_{ijkm} = \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} \\ + (abc)_{ijk} + p_m + q_{ijkm} + \epsilon_{ijkm}.$$

No further assumptions are involved in this decomposition, which is based on

an algebraic identity involving means and deviations over the array $\{Y_{ijkm}\}$. The explicit definitions and physical interpretations of the components of population model are delayed to Section 12 below. We note here that while the detail of the population model depends only on the experimental situation, the specific breakdown which we employ is determined by the design, since it will turn out that certain of the components of the model are estimable.

7. The statistical model; the function of randomization. We turn now to a consideration of the actual experimental observables. Let $x_{i^*j^*k^*f}$ denote the f th replicate observation obtained from selected factor combination $(i^*j^*k^*)$, where $f = 1, 2, \dots, n_{i^*j^*k^*}$. Since $x_{i^*j^*k^*f}$ is obtained from some one experimental unit, each $(i^*j^*k^*f)$ corresponds to some value of m , the experimental unit index. Against the background of the previous section we may regard the statistical effect of our experiment as giving a random (within the well-defined restrictions of the experimental design) sample, the $\{x_{i^*j^*k^*f}\}$, from the set of random variables $\{y_{ijkm}\}$; i.e., a restricted random sample of size $\sum_{i^*j^*k^*} n_{i^*j^*k^*}$ from the *ABCP* populations specified by the random variables $\{y_{ijkm}\}$.

It is appropriate to discuss here the function of randomization in this experimental design. Clearly, if we could observe the entire set $\{Y_{ijkm}\}$, we would know everything (empirically) possible about the experimental situation under consideration. Alternatively, if we could obtain observations on each member of the set $\{y_{ijkm}\}$, then only the technical errors $\{\epsilon_{ijkm}\}$ would be involved in our inferences about functions defined on elements of the set $\{Y_{ijkm}\}$. However, we are in general able to observe only a subset of the $\{y_{ijkm}\}$, and hence our inferences will be influenced by additional variabilities. The function of randomization is to attempt to control, in a statistical sense, these additional variabilities, and to enable us, perhaps, to obtain valid estimates of the uncertainties of inferences.

We incorporate the restrictions of the experimental design with the population model to obtain a statistical model for the observations, $\{x_{i^*j^*k^*f}\}$, in terms of parameters defined on elements of the set $\{Y_{ijkm}\}$ and of random variables which reflect (and define) the restrictions of the design. This statistical model has the advantage that it, together with the properties of its components, summarizes sufficiently all the relevant statistical knowledge and assumptions for the experiment. In addition certain results on linear estimation, variances of estimates, and expectations of analysis of variance mean squares may be derived by elementary algebraic operations using the statistical models. Furthermore there would be nothing more difficult than heavy algebra involved in obtaining more complex results, such as variances of mean squares, using the statistical model. It is to be expected, however, that more purely combinatorial arguments will shorten the process with regard to particular attributes (cf. Tukey [11] and Hooke [7]).

Full detail on the necessary additional notation and definitions needed to write down the statistical model is delayed till Section 12. At this point we note

that the statistical model takes the form

$$x_{i \cdot j \cdot k \cdot f} = \mu + a_i^* + b_j^* + c_k^* + (ab)_{i \cdot j}^* + (ac)_{i \cdot k}^* + (bc)_{j \cdot k}^* + (abc)_{i \cdot j \cdot k}^* + p_{i \cdot j \cdot k \cdot f}^* + q_{i \cdot j \cdot k \cdot f}^* + \epsilon_{i \cdot j \cdot k \cdot f}^*,$$

where, for example, $a_i^* = \sum_{i=1}^A \alpha_i^* a_i$, the a_i being parameters from the population model, the α_i^* being random variables which take on values zero or one with joint probability distribution specified by the experimental design. (In particular, for a_i^* the relevant item in the design is the random selection of a levels of factor A from the population of A levels.)

It is apparent from the subscripts in the above model that the last three components are mutually confounded, but their separation in the model is of importance because their statistical properties and experimental content are not alike.

The formal resemblance of the above statistical model (which may be appropriately called a definitional type model) to the usual "assumed linear models said to underly the analysis of variance" will be apparent and is not fortuitous. We note for emphasis that the model above depends only on the assumptions given in Section 6 above and not on any detailed knowledge or assumption concerning the mechanism (behaviour) of the experimental factors or units.

An extension of the application of the statistical model which we shall consider in this paper only very superficially would be to deduce certain elementary properties of the terms on the right-hand side (e.g., means and variances) and employ these with sufficient homogeneity and distributional assumptions to suggest a modified mathematical model which is more tractable from the point of view of "exact" distribution theory (cf. Scheffé [8]).

8. Succeeding sections. We invert the logical order of development by giving, in succeeding sections, results on expectations of analysis of variance mean squares (ems) in advance of definitions, notation, and derivations underlying these results. This is done because many who may be interested in the structure of these results will have much less concern with the detail of their derivation.

In Section 9 we deal with the case of proportional numbers (defined below) and on orthogonal⁵ analysis of variance based on weighted cell means; in Section 10 we consider the case of general numbers and a nonorthogonal analysis based on unweighted cell means; Section 11 deals with the special case of one factor. In addition to general formulae for expectations of mean squares, some questions of estimability of components of variation and of "proper error terms" are taken up.

In Section 12 we give details concerning the population and statistical models, explicit definitions of the components of variation, an example of the use of the statistical model in deriving ems,⁶ and discussion of various complements such

⁵ We use this term to refer to a decomposition in which the individual sums of squares sum to the so-called total sum of squares.

⁶ Expectations of mean squares.

as the physical interpretation of the parameters of the population model, relation of non-additivity to scale of observation, etc.

In Section 13 we describe briefly a more symmetric form for the results on ems which makes the extension to four or more factors very simple indeed. (This general pattern has been extended to include other experimental designs and its over-all structure will be described in later communications.)

Section 14 deals illustratively with problems of linear estimations, errors of estimates, and estimation of these errors, using the statistical model for these considerations.

Certain problems connected with the different roles of fixed and random factors and the need for functional structure analysis in the former are discussed by Wilk and Kempthorne [15] and will not be treated here.

9. The case of three factors, proportional numbers, no additivity assumptions.

We present in this section results on expectations of analysis of variance mean squares (henceforth referred to as ems) for the experimental situation and design given in Section 4, employing the conceptional framework described in Section 6, under the restriction that the number of observations in the subclasses fulfill the condition that

$$n_{i,j,k} = r u_i v_j w_k,$$

where r is the highest common factor of the $\{n_{i,j,k}\}$. Such a condition is often known as that of "proportional numbers."

Under these conditions an orthogonal analysis of variance, based on weighted means, exists. A case of "proportional numbers" can arise quite naturally when there are unequal numbers of observations corresponding to only one factor of classification.

The algebraic structure of the mean squares for such an analysis is well known; for example,

$$A^* = \frac{1}{(a-1)} \sum_{i,j,k,f} (x_{i\dots} - x_{\dots})^2$$

$$I_{AB}^* = \frac{1}{(a-1)(b-1)} \sum_{i,j,k,f} (x_{i,j\dots} - x_{i\dots} - x_{j\dots} + x_{\dots})^2,$$

where the usual dot convention is used to denote means.

We shall have use for the following notation:

$$U = \sum_i u_i; \quad V = \sum_j v_j; \quad W = \sum_k w_k;$$

$$U^* = \sum_i u_i^2/U^2; \quad V^* = \sum_j v_j^2/V^2; \quad W^* = \sum_k w_k^2/W^2.$$

(Note that for the case of equal numbers $U = a$, $V = b$, $W = c$, $U^* = 1/a$, $V^* = 1/b$, $W^* = 1/c$. Of course, in general, $U^* \leq 1$, $U^* \geq 1/a$.) Employing this notation, and recalling that f has range $1, 2, \dots, ru_i v_j w_k$, we obtain

$$A^* = \frac{1}{a-1} rVW \sum_i u_{i\cdot} (x_{i\cdot\cdot\cdot} - x_{\cdot\cdot\cdot})^2,$$

$$I_{AB}^* = \frac{1}{(a-1)(b-1)} rW \sum_{i,j} u_{i\cdot} v_{j\cdot} (x_{i\cdot j\cdot\cdot} - x_{i\cdot\cdot\cdot} - x_{\cdot\cdot j\cdot\cdot} + x_{\cdot\cdot\cdot\cdot})^2$$

General results on ems for this analysis are given in Table 1. The definitions of the components of variation⁷ which appear in the table are given in Section 12. For all the σ^2 's and Q^2 's with the exception of σ_a^2 the definition is such that they are a sum of squares of quantities divided by the number of linearly independent relations among these quantities. The subscript notation is intended to be suggestive; for example, σ_a^2 is a measure of dispersion of the population parameters $\{a_i\}$ which are the "main effects" of the levels of α ; σ_{ab}^2 reflects the dispersion of the population of interactions of levels of α with levels of β ; Q_{ap}^2 reflects the dispersion of the interactions of levels of A with experimental units; etc. (See Section 12 for further detail.) The definition of σ_a^2 requires a little comment. It is defined as

$$\sigma_a^2 = \frac{1}{ABC(P-1)} \sum_{ijkm} q_{ijkm}^2,$$

while the number of linearly independent relations among the set $\{q_{ijkm}\}$ is $(ABC-1)(P-1)$. The reason for this definition is partly because σ_a^2 appears in the ems for the residual and partly to simplify the formulae in Table 1. (Later, when we put Table 1 in a more symmetric form in Section 13, this disturbance will be eliminated.) The only distinction between the Q^2 's and the σ^2 's is that the former all reflect interactions of treatments with experimental units. The distinctive notation was employed to make this readily apparent in the table.

The results of Table 1 indicate that, in general, unbiased estimates of σ_a^2 , σ_b^2 , σ_c^2 , σ_{ab}^2 , etc., cannot be obtained from the analysis of variance mean squares if unit-treatment interactions are not negligible.⁸ The corresponding statement for the appropriate denominator in a test of significance criterion is complicated by possible ambiguity with respect to the null hypothesis of concern. But it is apparent that in a test of significance concerning, for example, the main effects of levels of α (see definitions of Section 12), we cannot in general find a "denominator" whose expectation is

$$E\left(A^* - rUVW \frac{(1-V^*)}{(a-1)} \sigma_a^2\right).$$

The question may arise as to whether it is in fact components such as σ_a^2

⁷ We refer to these quantities as "components of variation" rather than as "components of variance" to avoid possible ambiguity, since they are in fact measures of dispersion for the population of quantities on which they are defined, and are *not*, in the usual meaning of the word, variances of random variables.

⁸ This "bias" in the analysis of variance is the generalization of a similar result for a simpler situation, given by Wilk [12], [13].

TABLE 1

Expectations of mean squares for orthogonal analysis of variance

Mean squares	Expectation
A^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-U^*)}{(a-1)} \left\{ \left(V^* - \frac{1}{B} \right) \left(W^* - \frac{1}{C} \right) \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right] \right.$ $\left. + \left(W^* - \frac{1}{C} \right) \left[\sigma_{ac}^2 - \frac{1}{P} Q_{ac p}^2 \right] + \left(V^* - \frac{1}{B} \right) \left[\sigma_{ab}^2 - \frac{1}{P} Q_{ab p}^2 \right] + \left[\sigma_a^2 - \frac{1}{P} Q_{a p}^2 \right] \right\}$
B^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-V^*)}{(b-1)} \left\{ \left(U^* - \frac{1}{A} \right) \left(W^* - \frac{1}{C} \right) \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right] \right.$ $\left. + \left(W^* - \frac{1}{C} \right) \left[\sigma_{bc}^2 - \frac{1}{P} Q_{bc p}^2 \right] + \left(U^* - \frac{1}{A} \right) \left[\sigma_{ab}^2 - \frac{1}{P} Q_{ab p}^2 \right] + \left[\sigma_b^2 - \frac{1}{P} Q_{b p}^2 \right] \right\}$
C^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-W^*)}{(c-1)} \left\{ \left(U^* - \frac{1}{A} \right) \left(V^* - \frac{1}{B} \right) \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right] \right.$ $\left. + \left(V^* - \frac{1}{B} \right) \left[\sigma_{bc}^2 - \frac{1}{P} Q_{bc p}^2 \right] + \left(U^* - \frac{1}{A} \right) \left[\sigma_{ac}^2 - \frac{1}{P} Q_{ac p}^2 \right] + \left[\sigma_c^2 - \frac{1}{P} Q_{c p}^2 \right] \right\}$
I_{AB}^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-U^*)(1-V^*)}{(a-1)(b-1)}$ $\cdot \left\{ \left(W^* - \frac{1}{C} \right) \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right] + \left[\sigma_{ab}^2 - \frac{1}{P} Q_{ab p}^2 \right] \right\}$
I_{AC}^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-U^*)(1-W^*)}{(a-1)(c-1)}$ $\cdot \left\{ \left(V^* - \frac{1}{B} \right) \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right] + \left[\sigma_{ac}^2 - \frac{1}{P} Q_{ac p}^2 \right] \right\}$
I_{BC}^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-V^*)(1-W^*)}{(b-1)(c-1)}$ $\cdot \left\{ \left(U^* - \frac{1}{A} \right) \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right] + \left[\sigma_{bc}^2 - \frac{1}{P} Q_{bc p}^2 \right] \right\}$
I_{ABC}^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2 + rUVW \frac{(1-U^*)(1-V^*)(1-W^*)}{(a-1)(b-1)(c-1)} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2 \right]$
R^*	$\sigma^2 + \sigma_a^2 + \sigma_p^2$

which are of interest rather than the linear combination $[\sigma_a^2 - (1/P)Q_{ap}^2]$. The answer to this lies in an examination of the quantities $\{a_i\}$ whose dispersion make up σ_a^2 . By definition $a_i = Y_{i...} - Y_{...}$ and is thus the deviation of the average "true" response from level i of α , in combination with all levels of all

other factors and all experimental units, from the over-all average from all levels of all factors on all units. We refer to a_i as the main effect of the i th level of α . The difference between two such main effects $a_i - a_{i'}$ measures the difference (averaged over all levels of all other factors and all units) between operating at level i of α and level i' of α . On the other hand the combination

$$[\sigma_a^2 - (1/P)Q_{ap}^2]$$

is not always necessarily positive (though it will be in most cases of practical

TABLE 2

Error terms

Classification	Error terms
α	$V_A = R^* + \frac{(b-1)}{(1-V^*)} \left(V^* - \frac{1}{B} \right) (I_{AB}^* - R^*)$ $+ \frac{(c-1)}{(1-W^*)} \left(W^* - \frac{1}{C} \right) (I_{AC}^* - R^*)$ $- \frac{(b-1)(c-1)}{(1-V^*)(1-W^*)} \left(V^* - \frac{1}{B} \right) \left(W^* - \frac{1}{C} \right) (I_{ABC}^* - R^*)$
β	$V_B = R^* + \frac{(a-1)}{(1-U^*)} \left(U^* - \frac{1}{A} \right) (I_{AB}^* - R^*)$ $+ \frac{(c-1)}{(1-W^*)} \left(W^* - \frac{1}{C} \right) (I_{BC}^* - R^*)$ $- \frac{(a-1)(c-1)}{(1-U^*)(1-W^*)} \left(U^* - \frac{1}{A} \right) \left(W^* - \frac{1}{C} \right) (I_{ABC}^* - R^*)$
ϵ	$V_C = R^* + \frac{(a-1)}{(1-U^*)} \left(U^* - \frac{1}{A} \right) (I_{AC}^* - R^*)$ $+ \frac{(b-1)}{(1-V^*)} \left(V^* - \frac{1}{B} \right) (I_{BC}^* - R^*)$ $- \frac{(a-1)(b-1)}{(1-U^*)(1-V^*)} \left(U^* - \frac{1}{A} \right) \left(V^* - \frac{1}{B} \right) (I_{ABC}^* - R^*)$
$\alpha \times \beta$	$V_{AB} = R^* + \frac{(c-1)}{(1-W^*)} \left(W^* - \frac{1}{C} \right) (I_{ABC}^* - R^*)$
$\alpha \times \epsilon$	$V_{AC} = R^* + \frac{(b-1)}{(1-V^*)} \left(V^* - \frac{1}{B} \right) (I_{ABC}^* - R^*)$
$\beta \times \epsilon$	$V_{BC} = R^* + \frac{(a-1)}{(1-U^*)} \left(U^* - \frac{1}{A} \right) (I_{ABC}^* - R^*)$
$\alpha \times \beta \times \epsilon$	$V_{ABC} = R^*$

interest) and hence is not a measure of dispersion of *any* quantities defined on the basic population of true responses $\{Y_{ijkm}\}$.

Two factors tend to decrease the importance of this "bias" in the analysis of variance due to interactions of treatments with experimental units. First, the quantity confounded with the component of variation of interest enters in the ems with coefficient $1/P$. Thus if P , the number of experimental units is large, then the effect of the confounded term will usually be small. The origin of the confounded term is the negative correlation induced on observed responses for a given treatment combination owing to the random assignment of units from a finite population. As the size of the population of units increases, this correlation goes to zero. Secondly, each Q^2 quantity represents a higher order interaction term than the component with which it is associated, and it is often true that the higher the order of the interaction the smaller it will be. The size of unit treatment interactions depends somewhat independently on two considerations, namely, the scale of measurement of the responses and the heterogeneity of the experimental units. Of course, homogeneous experimental units will mean additivity of units and treatments on any scale.

Under the assumption that all unit-treatment interactions are zero (i.e., that $q_{ijkm} = 0$) so-called proper error terms would exist. Table 2 lists error terms for each classification of the design. The bias in using these error terms when unit-treatment interactions are not negligible is exemplified by

$$[-rUVW(1 - U^*) / (a - 1)(1/P)Q_{sp}^2],$$

which is the bias in using V_A as an error term for α main effects.

As we shall see in a later section, the device of randomization is fully effective in allowing unbiased linear estimation of treatment effects. But essentially unbiased error terms will be obtainable from the analysis of variance, in general, only when the experimental units are not too heterogeneous or the size, P , of the population of units is large, or the scale is such that units and treatment combinations are additive (in the sense that their interactions on that scale are zero.) There does not appear to be any simple statistical method to overcome this confounding which is due to the "fractional replication" which is imposed by the restriction that each unit can be "used only once."

We close this section with a discussion of three special cases which have been given much attention in the past. For simplicity we reduce our consideration to those involving two factors, α and β , putting $C = c = 1$, and shall take P as "very large." The cases we detail are the so-called "fixed," "mixed," and "random" model situations. The results on ems are then those of Table 3: $\sigma_0^2 = \sigma_p^2 + \sigma_q^2$.

The following points from Table 3 are worthy of note: If the numbers of observations in each "cell" are equal, then $U^* = 1/a$ and $V^* = 1/b$ and then the component $\sigma_{\alpha\beta}^2$ vanishes from $E(A^*)$ and from $E(B^*)$ for the fixed case; and from $E(B^*)$ in the mixed case, where α is the "fixed" factor, but not from $E(A^*)$, where β is the random factor. If the numbers are proportional and not equal

TABLE 3
Ems for special cases of a two-factor experiment

Mean square	1. Fixed: $A = a, B = b$	2. Mixed: $A = a, B \gg b$	3. Random: $A \gg a, B \gg b$
A^*	$\sigma_0^2 + rUV \frac{(1 - U^*)}{(a - 1)}$ $\cdot \left[\left(V^* - \frac{1}{b} \right) \sigma_{ab}^2 + \sigma_a^2 \right]$	$\sigma_0^2 + rUV \frac{(1 - U^*)}{(a - 1)}$ $\cdot [V^* \sigma_{ab}^2 + \sigma_a^2]$	Same as 2.
B^*	$\sigma_0^2 + rUV \frac{(1 - V^*)}{(b - 1)}$ $\cdot \left[\left(U^* - \frac{1}{a} \right) \sigma_{ab}^2 + \sigma_b^2 \right]$	Same as 1.	$\sigma_0^2 + rUV \frac{(1 - V^*)}{(b - 1)}$ $\cdot [U^* \sigma_{ab}^2 + \sigma_b^2]$
I_{AB}^*	$\sigma_0^2 + rUV \frac{(1 - U^*)}{(a - 1)}$ $\cdot \frac{(1 - V^*)}{(b - 1)} \sigma_{ab}^2$	Same as 1.	Same as 1.
R^*	σ_0^2		

then, even for these special cases, we do not have simple comparability of the orthogonal analysis of variance mean squares, as has been pointed out by Smith [9]. The fact that, for the case of equal numbers in the mixed case, the component due to interaction remains associated with the fixed factor but not with the random factor is due, loosely speaking, to our having information on each observed "random" factor level in combination with every fixed factor level in our population; but for each fixed factor level we have only a random selection from the possible random factor levels. The crucial point is that for the case α fixed, β random, σ_a^2 reflects the dispersion of effects of levels of α averaged over all levels of β and similarly for σ_b^2 ; and while every level of α is used in the experiment, only a sample of levels of β are studied.

10. The case of three factors, general numbers, no additivity assumptions. In the event that no restrictions are placed on the numbers $n_{i,j,k}$, except that they be non-zero, an orthogonal analysis of variance, in which the various sums of squares all have a meaningful relationship to the experimental situation, for a multiple factor experiment will not, in general, exist. One can, however, make an analysis of variance based on cell means. The algebraic structure of such an analysis is exemplified as follows: Let A^{**} be the mean square associated with α main effects in this analysis, and let

$$\bar{x}_{i\dots} = \frac{1}{bc} \sum_{j,k} x_{i,j,k},$$

and

$$\bar{x}_{\dots} = \frac{1}{abc} \sum_{i,j,k} x_{i,j,k};$$

then,

$$A^{**} = bc \sum_i (\bar{x}_{i\dots} - \bar{x}_{\dots})^2 / (a - 1).$$

The table is completed by a line for residual mean square, R^{**} , which reflects "within cell" deviations and is in fact identical with R^* of Section 9. This analysis is not orthogonal in the sense that the individual sums of squares will not, in general, sum to the so-called total sum of squares,

TABLE 4
Expected mean squares for non-orthogonal analysis of variance

Mean square	Expectation of mean square
A^{**}	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \frac{(B-b)(C-c)}{B} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right]$ $+ \frac{(C-c)}{C} b \left[\sigma_{ac}^2 - \frac{1}{P} Q_{acp}^2 \right] + \frac{(B-b)}{B} c \left[\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2 \right] + bc \left[\sigma_a^2 - \frac{1}{P} Q_{ap}^2 \right]$
B^{**}	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \frac{(A-a)(C-c)}{A} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right]$ $+ \frac{(C-c)}{C} a \left[\sigma_{bc}^2 - \frac{1}{P} Q_{bcp}^2 \right] + \frac{(A-a)}{A} c \left[\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2 \right] + ac \left[\sigma_b^2 - \frac{1}{P} Q_{bp}^2 \right]$
C^{**}	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \frac{(A-a)(B-b)}{A} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right]$ $+ \frac{(B-b)}{B} a \left[\sigma_{bc}^2 - \frac{1}{P} Q_{bcp}^2 \right] + \frac{(A-a)}{A} b \left[\sigma_{ac}^2 - \frac{1}{P} Q_{acp}^2 \right] + ab \left[\sigma_c^2 - \frac{1}{P} Q_{cp}^2 \right]$
I_{AB}^*	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \frac{(C-c)}{C} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right] + c \left[\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2 \right]$
I_{AC}^{**}	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \frac{(B-b)}{B} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right] + b \left[\sigma_{ac}^2 - \frac{1}{P} Q_{acp}^2 \right]$
I_{BC}^{**}	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \frac{(A-a)}{A} \left[\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right] + a \left[\sigma_{bc}^2 - \frac{1}{P} Q_{bcp}^2 \right]$
I_{ABC}^{**}	$\frac{1}{n^*} (\sigma^2 + \sigma_q^2 + \sigma_p^2) + \left(\sigma_{abc}^2 - \frac{1}{P} Q_{abcp}^2 \right)$
R^{**}	$(\sigma^2 + \sigma_q^2 + \sigma_p^2)$

$$\sum_{i,j,k,f} (x_{i,j,k,f} - x_{\dots})^2.$$

The only exception to this last statement (when dealing with two or more factors) is when the numbers $n_{i,j,k}$ are all equal.

The statistical model appropriate here is identical with that used for Section 9 and is developed in Section 12. Table 4 gives the ems for this analysis, with no additivity assumptions. We employ the notation

$$\frac{1}{n^*} = \frac{1}{abc} \sum_{i,j,k} \left(\frac{1}{n_{i,j,k}} \right) = \text{average value of elements of the set } \left\{ \frac{1}{n_{i,j,k}} \right\}.$$

Definitions of components of variation are the same as in Section 9 and are detailed in Section 12.

The advantage attached to this analysis of variance is the simple structure of the expectations of mean squares, as opposed to the very complex relations exhibited in Table 1. In fact, if all mean squares in Table 4, except R^{**} , are adjusted by multiplying by n^* then, speaking rather loosely, this analysis may be superficially interpreted in a similar way to an analysis for a case with equal numbers in the cells. (For equal numbers, n^* becomes simply r .)

The discussion given in the preceding section in connection with difficulties when unit-treatment interactions are not negligible applies also to the non-orthogonal analysis. If unit treatment interactions are negligible, then one can obtain from linear combinations of the mean squares of the non-orthogonal analysis unbiased estimates of the various components of variation of interest. For example, with negligible unit treatment interactions an unbiased estimate of σ_{ab}^2 is given by

$$\frac{1}{c} \left[I_{Ab}^{**} - \frac{(C-c)}{C} \left(I_{Abc}^{**} - \frac{I}{n^*} R^{**} \right) - \frac{1}{n^*} R^{**} \right].$$

The relation of this to the selection of appropriate "error terms" to serve as denominators in F -type comparisons of mean squares will be apparent. The relation to the estimation of variances of linear estimates is no less immediate and is dealt with explicitly in Section 14.

If one has a situation involving proportional but unequal numbers, the question arises whether one should employ the orthogonal analysis based on weighted means or the non-orthogonal analysis based on unweighted means. In the present state of knowledge it appears to be a matter of taste, convenience, and opinion as to which analysis is more advantageous. (Some recent relevant references on this point are Cox [2] and Tukey [11]).

The non-orthogonal analysis has the advantages of wider generality, easier computations, simpler terms and more direct connection with the estimation of linear contrasts among treatment effects. Furthermore, speaking very loosely, the non-centrality enters into the mean squares of the non-orthogonal analysis in a more easily appreciated and more symmetric fashion than for the orthogonal analysis.

The questions of efficiency of estimation of components of variation and of sensitivity of significance, as regards these two analyses, are still open.

11. One factor, general numbers. The case of two factors may be obtained as a special case of the three-factor development by putting $C = c = 1$. We will not deal with it explicitly. The case of one factor can be obtained by putting $B = b = 1$, in addition. Because of some peculiarities in this situation we give some brief discussion below.

The one factor case corresponds to the within and between analysis of variance and has the associated property that an orthogonal analysis always exists whatever the numbers of observations corresponding to the various levels tested. On the other hand one still has the choice as to whether to analyze weighted or unweighted means of observation corresponding to the levels tested.

The residual mean square is the same for both analyses. For the proportional analysis

$$A^* = \frac{r}{(a-1)} \sum_{i*} u_{i*} (x_{i*} - \bar{x}_{..})^2,$$

where

$$x_{i*} = \frac{1}{n_{i*}} \sum_f x_{i*f} \quad \text{and} \quad \bar{x}_{..} = \frac{1}{\sum_i n_{i*}} \sum_{i,f} x_{i*f};$$

$$E(A^*) = \sigma_0^2 + rU \frac{(1-U^*)}{(a-1)} \left[\sigma_a^2 - \frac{1}{P} Q_{ap}^2 \right],$$

where $\sigma_0^2 = \sigma^2 + \sigma_p^2 + \sigma_q^2$, $n_{i*} = ru_{i*}$, $U = \sum_i u_{i*}$, $U^* = \sum_i u_{i*}^2 / U^2$. For the non-orthogonal analysis,

$$A^{**} = \frac{1}{(a-1)} \sum_{i*} (x_{i*} - \bar{x}_{..})^2, \quad \text{where } \bar{x}_{..} = \frac{1}{a} \sum_{i*} x_{i*};$$

$$E(A^{**}) = \frac{1}{n^*} \sigma_0^2 + \left[\sigma_a^2 - \frac{1}{P} Q_{ap}^2 \right], \quad \text{where } \frac{1}{n^*} = \frac{1}{a} \sum_{i*} \frac{1}{n_{i*}}.$$

Thus, in the non-orthogonal analysis of variance equal weight is given to each observed level of the factor. In the case of a single factor there does not appear, offhand at least, to be any basis to suggest that one analysis will be, in general, superior to the other.

12. Derivation of models and ems. Our attention is directed in this section to the following main items: (i) definitions and physical interpretations of the parameters of the population model; (ii) the explicit development of a formal statistical model for the observations; (iii) definitions for the various components of variation; (iv) illustration of the use of the statistical model in the derivation of ems.

In Section 6 we gave a conceptional framework for the analysis of the general three-factor completely randomized experiment. This specified as the background

population a set of ABCP (unknown) numbers, $\{Y_{ijkm}\}$, the "true" or "typical" responses. A useful and meaningful representation of these is the one implied by the definitions

$$\begin{aligned}\mu &= Y_{\dots}, \\ a_i &= Y_{i\dots} - \mu, \\ b_j &= Y_{.j\dots} - \mu, \\ c_k &= Y_{\dots k} - \mu, \\ (ab)_{ij} &= Y_{ij\dots} - Y_{i\dots} - Y_{.j\dots} + \mu, \\ (ac)_{ik} &= Y_{i.k\dots} - Y_{i\dots} - Y_{\dots k} + \mu, \\ (bc)_{jk} &= Y_{.jk\dots} - Y_{.j\dots} - Y_{\dots k} + \mu, \\ (abc)_{ijk} &= Y_{ijk\dots} - Y_{ij\dots} - Y_{i.k\dots} - Y_{.jk\dots} + Y_{i\dots} + Y_{.j\dots} + Y_{\dots k} - (bc)_{jk}, \\ p_m &= Y_{\dots m} - \mu, \\ q_{ijkm} &= Y_{ijkm} - Y_{ijk\dots} - Y_{\dots m} + \mu.\end{aligned}$$

It is easy to check that the sum of all the components defined above is identically equal to Y_{ijkm} . We have now

$$(1 + A + B + C + AB + AC + BC + ABC + P + ABCP)$$

quantities, in place of our original $ABCP$, but the following properties indicate the dependencies:

$$\begin{aligned}0 &= \sum_i a_i = \sum_j b_j = \sum_k c_k = \sum_{ij} (ab)_{ij} = \sum_{ik} (ac)_{ik} = \sum_{jk} (bc)_{jk} \\ &= \sum_{ijk} (abc)_{ijk} = \sum_m p_m = \sum_{ijk} q_{ijkm} = \sum_m q_{ijkm}.\end{aligned}$$

These relations follow by definition of the parameters and *not* by assumption.

The quantities defined above can be given physical interpretation. We shall do this for representative cases:

μ is the "true" over-all conceptual response if all treatment combinations were applied to all experimental units.

a_i is the difference between the mean of the "true" responses if all treatments consisting of the i th level of α in combination with every level of β and every level of γ were applied to all experimental units, and μ ; we refer to a_i as the main effect or simply the effect of the i th level of α . It should be noted that $a_i - a_{i'}$ is the difference between the responses due to level i of α and level i' of α averaged over all levels of other factors and all units.

$(ab)_{ij}$ is the difference between the effect of the j th level of β in combination with the i th level of α and the main effect of the j th level of β . (The symmetry between α and β is obvious from the definition of $(ab)_{ij}$.) We call $(ab)_{ij}$ the

interaction effect, or simply the interaction, of the i th level of α and the j th level of β .

p_m measures the difference between the mean response from all combinations of levels of α , β , and γ on unit m compared to μ . Thus the p_m measure the (average) variability of units with respect to the treatment combinations. Because of the direction of our interest, we refer to the p_m as the additive unit errors or simply as unit errors.

q_{ijkm} is similarly seen to represent the interaction of treatment combination (ijk) with unit m : and we refer to the q_{ijkm} as interactive unit errors or unit-treatment interactions.

A number of items deserve explicit mention even though some have been indicated in the literature, specifically by Yates [16, 17, 18] and others. (1) The definitions of effects and interactions are relative to a given scale of measurement of response. Transformations of the scale would lead to radically different effects and interactions associated with the treatment population. To speak of the main effect of level 2 of factor α as being large is meaningless unless a particular scale of response is implicit. Similarly the entire concept of interaction has meaning only relative to a given scale. Two factors can, with no contradiction, have negligible interactions on one scale and large interactions on another scale. (2) For a given scale of measurement of response, the definition (and interpretation) of, say, the effect of the i th level of α depends not only on all other levels of α included in the experimental population but also on all levels of all other factors as well as on the relevant population of experimental units. The generalization to other effects and interactions is immediate. It is of interest and importance to note that the difference of the effects, say $a_i - a'_i$, of two levels of α becomes independent of other levels of α under consideration but remains entirely dependent on the levels of the *other* factors and the population of experimental units. (3) If we have a scale of observation such that, for instance, all interactions $(ab)_{ij}$ are negligible, or α and β are additive, then the difference $a_i - a'_i$ becomes independent of which levels of β are included in the study. This points up the enormous simplification in the summarization of relevant information and in understanding of the situation which is effected when we can operate on a scale in which interactions may be neglected. (4) If the levels of factor are essentially identical in terms of their influence on response, then, on any scale, interactions with that factor will be negligible. Similarly, if experimental units are fairly homogeneous, then one would expect that, for most scales of observations, the variability of units would be largely described by the unit errors, p_m , and the unit-treatment interactions would be negligible. (5) The "reparametrization" of the population of "true responses" to effects, interactions, and unit errors focuses our attention on summary properties of the experimental situation. This has the advantages that (i) the analysis of variance mean squares are interpretable in terms of these parameters, which have a physical interpretation; (ii) knowledge of certain of the parameters is often essentially the information we desire from the experiment; (iii) by means of the decomposition given by the population model it is often simpler to appreciate

and evaluate assumptions which may be implicit or explicit in a particular procedure or inference.

We turn our attention now to the development of a formal usable statistical model, for the actual experimental observations, in terms of the parameters defined above. We recall that our experiment could be regarded as giving us a random (within the restrictions of the design) sample, the $\{x_{i^*j^*k^*f}\}$, of size $\sum_{i^*j^*k^*f} n_{i^*j^*k^*f}$, from the set of *ABCP* populations represented by the random variables $\{y_{ijkm}\}$, where $y_{ijkm} = Y_{ijkm} + \epsilon_{ijkm}$. To write an explicit model for the $x_{i^*j^*k^*f}$ it is useful and convenient to employ certain "dummy" random variables which we now proceed to define.

Let $\alpha_i^{i^*} = 1$ if selected level i^* of factor corresponds to level i in the population of levels of α ;

= 0 otherwise.

Thus, if the 2nd selected level of factor α corresponds to the 5th population level of factor α , then $\alpha_5^2 = 1$.

Similarly we define the sets $\{\beta_j^{j^*}\}$ and $\{\gamma_k^{k^*}\}$.

Because of the specification of random selection these quantities are random variables some of whose distributional properties are easily written down. For example; (1) The $\{\alpha_i^{i^*}\}$, $\{\beta_j^{j^*}\}$, $\{\gamma_k^{k^*}\}$ are groupwise statistically independent; (2) $\Pr(\alpha_i^{i^*} = 1) = 1/A$; (3) $P(\alpha_i^{i^*} \alpha_{i'}^{i'^*} = 0) = 1, i \neq i'$; (4) $P(\alpha_i^{i^*} \alpha_{i'}^{i'^*} = 1) = (1/A(A-1))$, $i^* \neq i'^*$, $i \neq i'$; (5) $P(\beta_j^{j^*} = 0) = (B-1)/B$; etc. We note that the α 's, β 's, and γ 's are associated with the random selection of the factor levels to be tested.

We turn now to the specification of association of selected treatment combinations with experimental units. To this end we define

$\delta_m^{i^*j^*k^*f} = 1$ if the f th replicate of selected treatment combination $(i^*j^*k^*)$ is tested on unit m of the population of experimental units;
= 0, otherwise.

In view of the random selection of units for test and the randomization of treatment combinations to experimental units, it follows that the $\{\delta_m^{i^*j^*k^*f}\}$ are random variables with the following properties: (1) They are statistically independent of the α 's, β 's, and γ 's defined above; (2) $P(\delta_m^{i^*j^*k^*f} = 1) = 1/P$; (3) $P(\delta_m^{i^*j^*k^*f} \delta_{m'}^{i'^*j'^*k'^*f'} = 0) = 1, (i^*j^*k^*f) \neq (i'^*j'^*k'^*f')$; (4) $P(\delta_m^{i^*j^*k^*f} \delta_{m'}^{i'^*j'^*k'^*f'} = 0) = 1, m \neq m'$; (5) $P(\delta_m^{i^*j^*k^*f} \delta_{m'}^{i'^*j'^*k'^*f'} = 1) = (1/P(P-1))$, $(i^*j^*k^*f) \neq (i'^*j'^*k'^*f')$, $m \neq m'$; etc.

It is now simple to write an explicit model for the observations $x_{i^*j^*k^*f}$, as follows:

$$\begin{aligned} x_{i^*j^*k^*f} = & \mu + \sum_i \alpha_i^{i^*} a_i + \sum_j \beta_j^{j^*} b_j + \sum_k \gamma_k^{k^*} c_k + \sum_{ij} \alpha_i^{i^*} \beta_j^{j^*} (ab)_{ij} \\ & + \sum_{ik} \alpha_i^{i^*} \gamma_k^{k^*} (ac)_{ik} + \sum_{jk} \beta_j^{j^*} \gamma_k^{k^*} (bc)_{jk} + \sum_{ijk} \alpha_i^{i^*} \beta_j^{j^*} \gamma_k^{k^*} (abc)_{ijk} \\ & + \sum_m \delta_m^{i^*j^*k^*f} p_m + \sum_{ijkm} \alpha_i^{i^*} \beta_j^{j^*} \gamma_k^{k^*} \delta_m^{i^*j^*k^*f} (q_{ijkm} + \epsilon_{ijkm}). \end{aligned}$$

The correspondence of terms between this and what is given in Section 7 will be apparent on inspection.

From the point of view of our development, the random variables in this model are the α 's, β 's, γ 's, and δ 's, which take on the values 0 and 1 with probabilities specified by the experimental design and procedure, and the ϵ 's. All other quantities are regarded as fixed, unknown parameters defined on the array of "true" responses $\{Y_{ijkm}\}$.

This model, together with the properties of its components, contains all the implications of our procedures of random selection and allocation, as well as all assumptions we have made in the conceptual frame of reference for the analysis. It is in a sense "sufficient" for the general experimental situation and design, together with the additional assumptions which we made explicit in Section 6. This model can therefore be employed quite formally in any statistical manipulation or evaluations of the experiment, without reference to any other features.

The complexity of the model is only in its initial appearance. It is easy to handle in algebraic manipulations and, in particular, makes into an elementary algebraic operation the evaluation of expectations of various functions of the observations.

Toward the end of this section we shall illustrate how the statistical model is employed in evaluation of expectations of analysis of variance mean squares. Before doing this we give the explicit definitions of the components of variation, which have appeared in the ems in previous sections, in terms of the components of the population model. These are as follows:

$$\begin{aligned}\sigma_a^2 &= \frac{1}{A-1} \sum_i a_i^2; & \sigma_b^2 &= \frac{1}{B-1} \sum_j b_j^2; & \sigma_c^2 &= \frac{1}{C-1} \sum_k c_k^2; \\ \sigma_{ab}^2 &= \frac{1}{(A-1)(B-1)} \sum_{ij} (ab)_{ij}^2; & \sigma_{ac}^2 &= \frac{1}{(A-1)(C-1)} \sum_{ik} (ac)_{ik}^2; \\ \sigma_{bc}^2 &= \frac{1}{(B-1)(C-1)} \sum_{jk} (bc)_{jk}^2; & \sigma_{abc}^2 &= \frac{1}{(A-1)(B-1)(C-1)} \sum_{ijk} (abc)_{ijk}^2; \\ \sigma_p^2 &= \frac{1}{(P-1)} \sum_m p_m^2; & \sigma_q^2 &= \frac{1}{ABC(P-1)} \sum_{ijkm} q_{ijkm}^2; & \sigma^2 &= E(\epsilon_{ijkm}^2); \\ Q_{ap}^2 &= \frac{1}{(A-1)(P-1)} \sum_{im} q_{i..m}^2; & Q_{bp}^2 &= \frac{1}{(B-1)(P-1)} \sum_{jm} q_{.j.m}^2; \\ Q_{cp}^2 &= \frac{1}{(C-1)(P-1)} \sum_{km} q_{..km}^2; \\ Q_{abp}^2 &= \frac{1}{(A-1)(B-1)(P-1)} \sum_{ijm} (q_{ij..m} - q_{i..m} - q_{.j.m})^2; \\ Q_{acp}^2 &= \frac{1}{(A-1)(C-1)(P-1)} \sum_{ikm} (q_{i.km} - q_{i..m} - q_{..km})^2;\end{aligned}$$

$$Q_{bcp}^2 = \frac{1}{(B-1)(C-1)(P-1)} \sum_{ijk} (q_{ijk} - q_{i \cdot \cdot m} - q_{\cdot j \cdot m} - q_{\cdot \cdot km})^2;$$

$$Q_{abcp}^2 = \frac{1}{(A-1)(B-1)(C-1)(P-1)} \sum_{ijklm} (q_{ijklm} - q_{ij \cdot \cdot m} - q_{i \cdot km} + q_{i \cdot \cdot m} - q_{\cdot jkm} + q_{\cdot j \cdot m} + q_{\cdot \cdot km})^2.$$

With the exception of σ_q^2 the definition of these components is according to the scheme

$$\frac{(\text{sum of squares of quantities})}{(\text{no. of quantities} - \text{no. of linear dependencies})}.$$

While there is no doubt that the σ_a^2 , σ_b^2 , etc., reflect the variability of the populations $\{a_i\}$, $\{b_j\}$, etc., some further justification for the method of choice of divisors is in order. An important (and perhaps sufficient) justification is that such a method of definition simplifies the appearance of the ems and the variances of certain linear estimates. For further insight we might argue that the measure of dispersion wanted for, say, the $\{a_i\}$ is essentially that for the $\{Y_{i \cdot \cdot}\}$, a fundamental measure of the dispersion of which is one of Gini's mean differences, namely, the average of squares of differences between pairs from the population. For the case of the $\{Y_{i \cdot \cdot}\}$ this is

$$\begin{aligned} G_a &= \frac{1}{A(A-1)} \sum_{i \neq i'} (Y_{i \cdot \cdot} - Y_{i' \cdot \cdot})^2 \\ &= \frac{2}{A-1} \sum_i (Y_{i \cdot \cdot} - Y_{\cdot \cdot \cdot})^2 \\ &= 2\sigma_a^2. \end{aligned}$$

(The factor 2 arises because each pair, in inverted order, appears twice.) The same argument applies to σ_b^2 , σ_c^2 , and σ_p^2 . For the case of a measure of dispersion of, say, the $\{(ab)_{ij}\}$, we might argue that this should reflect the magnitude of interactions in the two-way array $\{Y_{ij \cdot \cdot}\}$, a fundamental measure of which is a mean square "double difference"

$$G_{ab} = \frac{1}{AB(A-1)(B-1)} \sum_{\substack{i \neq i' \\ j \neq j'}} [(Y_{ij \cdot \cdot} - Y_{i'j \cdot \cdot}) - (Y_{ij' \cdot \cdot} - Y_{i'j' \cdot \cdot})]^2.$$

Now the quantity in square brackets is identical with

$$(ab)_{ij} - (ab)_{ij'} - (ab)_{i'j} + (ab)_{i'j'},$$

and remembering that $\sum_i (ab)_{ij} = \sum_j (ab)_{ij} = 0$, and hence that

$$\sum_{ij} (ab)_{ij}^2 = - \sum_{\substack{i \neq i' \\ j \neq j'}} (ab)_{ij} (ab)_{ij'} = - \sum_{\substack{i \neq i' \\ j}} (ab)_{ij} (ab)_{i'j} = \sum_{\substack{i \neq i' \\ j \neq j'}} (ab)_{ij} (ab)_{i'j'},$$

it is easy to find that

$$G_{ab} = 4\sigma_{ab}^2.$$

(Again, the factor 4 arises because essentially the same quantity is permitted to appear four times.) The same argument applies to σ_{ac}^2 , σ_{bc}^2 , Q_{ap}^2 , Q_{bp}^2 , and Q_{cp}^2 , and can be extended in the obvious way to σ_{abc}^2 , Q_{abp}^2 , Q_{acp}^2 , Q_{bcp}^2 , and Q_{abcp}^2 .

The structure of the Q^2 quantities is, from their definition and that of q_{ijkm} , such that they reflect interactions of treatment factors with experimental units. For example

$$Q_{ap}^2 = \frac{1}{(A-1)(B-1)} \sum_{im} (Y_{i...m} - Y_{i...} - Y_{...m} + Y_{....})^2,$$

which reflects the interactions of levels of α with experimental units. In view of the role which the unit treatment interaction components of variation play in the ems, it was felt that a distinctive notation for them would be worth while. So far as their formal definitions are concerned there are no distinctions between the σ^2 's (except σ_a^2) and the Q^2 's.

The essential reasons for the definition of σ_a^2 which was used are that σ_a^2 appears in the expectation of the residual mean square and that such a definition shortens some of the formulae. It is easily checked that

$$\begin{aligned} \sigma_a^2 = & \frac{(A-1)}{A} Q_{ap}^2 + \frac{(B-1)}{B} Q_{bp}^2 + \frac{(C-1)}{C} Q_{cp}^2 + \frac{(A-1)(B-1)}{AB} Q_{abp}^2 \\ & + \frac{(A-1)(C-1)}{AC} Q_{acp}^2 + \frac{(B-1)(C-1)}{BC} Q_{bcp}^2 \\ & + \frac{(A-1)(B-1)(C-1)}{ABC} Q_{abcp}^2. \end{aligned}$$

We proceed now to show how to use the statistical model in deriving the expectation of the α mean square, $A^* = 1/(a-1)A'$, for the case of proportional numbers (Section 9). This will illustrate the basis for the results given in previous sections.

We have

$$A' = \sum_{i,j,k,\dots} n_{i,j,k,\dots} (x_{i,j,k,\dots} - x_{....})^2 = rVW \sum_{i^*} u_{i^*} (x_{i^*} - x_{....})^2.$$

The statistical model can now be substituted into this expression, and determining the expectation becomes a purely algebraic operation when one uses freely the fact that the expectation of a sum is the sum of the expectations. Thus,

$$\begin{aligned} A' = & rVW \sum_{i^*} u_{i^*} [a_{i^*}^* - a^* + (ab)_{i^*}^* - (ab)^* + (ac)_{i^*}^* - (ac)^* \\ & + (abc)_{i^*}^* - (abc)^* + p_{i^*}^* - p^* + q_{i^*}^* - q^* + \epsilon_{i^*}^* - \epsilon^*]^2, \end{aligned}$$

where

$$a_{i^*}^* = \sum_i \alpha_i^* a_i; \quad a^* = \frac{1}{U} \sum_{i^*} u_{i^*} a_{i^*}^*;$$

$$(ab)_{i^*}^* = \frac{1}{V} \sum_{j^*} v_{j^*} (ab)_{i^* j^*}^* = \frac{1}{V} \sum_{j^* i^*} v_{j^*} \alpha_{i^*}^* \beta_{j^*}^* (ab)_{ij};$$

$$(ab)_{i^*}^* = \frac{1}{UV} \sum_{j^*} u_{i^*} v_{j^*} (ab)_{i^* j^*}^*;$$

$$p_{i^*}^* = \frac{1}{ru_{i^*} VW} \sum_{j^* k^* f} \delta_m^{i^* j^* k^* f} p_m;$$

$$q_{i^*}^* = \frac{1}{ru_{i^*} VW} \sum_{j^* k^* f} \alpha_{i^*}^* \beta_{j^*}^* \gamma_{k^*}^* \delta_m^{i^* j^* k^* f} q_{ijkm};$$

etc.

It is easy to check that unlike terms in the above expression are uncorrelated. For example,

$$E(a_{i^*}^* p_{i^*}^* \dots) = E\left(\sum_i \alpha_i^* a_i\right) \left(\sum_{j^* k^* f} \delta_m^{i^* j^* k^* f} p_m\right) = \left[\sum_i a_i E(\alpha_i^*)\right] \left[\sum_{j^* k^* f} p_m E(\delta_m^{i^* j^* k^* f})\right],$$

since the α 's and δ 's are independent. But $E(\alpha_i^*) = 1/A$ for all i and i^* ;

$$E(\delta_m^{i^* j^* k^* f}) = 1/P,$$

for all m, i^*, j^*, k^* , and f , and $\sum_i a_i = \sum_m p_m = 0$. Similarly, the expectation of all other cross-product terms may be shown to be zero.

(In the event that, for example, α is fixed, i.e., $A = a$, one will in general not renumber the levels at random, so that in our notation i^* and i would be the same index in making the formal correspondence. As we mentioned elsewhere, for symmetric functions of the observations (in our sample involving all levels of α) no difficulty arises. In the section on linear estimation which involves non-symmetric functions we shall give an extended notation. For the present, if we use the convention that when $A = a$, i^* and i are the same index, then, for example, $\sum_i \alpha_i^3 a_i = a_3$, since $\alpha_i^i = 1$ with probability 1 and $\alpha_i^{i'} = 0$, $i \neq i'$ with probability 1, using our convention. Then $a_{i^*}^*$ would become a_i , a constant, and since $E(p_{i^*}^* \dots) = 0$, the above result and its analogues remain true.)

Hence

$$E(A') = rVW \sum_{i^*} u_{i^*} E\{[a_{i^*}^* - a^*]^2 + [(ab)_{i^*}^* - (ab)^*]^2 + [(ac)_{i^*}^* - (ac)^*]^2 + [(abc)_{i^*}^* - (abc)^*]^2 + [p_{i^*}^* - p^*]^2 + [q_{i^*}^* - q^*]^2 + [\epsilon_{i^*}^* - \epsilon^*]^2\}.$$

Now,

$$\begin{aligned} \sum_{i^*} u_{i^*} E(a_{i^*}^* - a^*)^2 &= E\left\{\sum_{i^*} u_{i^*} a_{i^*}^{*2} - \frac{1}{U} \left(\sum_{i^*} u_{i^*} a_{i^*}^*\right)^2\right\} \\ &= E\left\{\sum_{i^*} u_{i^*} \left(\sum_i \alpha_i^{i^*} a_i\right)^2 - \frac{1}{U} \left(\sum_{i^*} u_{i^*} \sum_i \alpha_i^{i^*} a_i\right)^2\right\} \\ &= E\left\{\sum_{i^* j} u_{i^*} \alpha_i^{i^*} a_i^2 - \frac{1}{U} \sum_{i^*} u_{i^*}^2 \alpha_i^{i^*} a_i^2 - \frac{1}{U} \sum_{\substack{i^* \neq i' \\ i \neq i'}} u_{i^*} u_{i'} \alpha_i^{i^*} \alpha_{i'}^{i'} a_i a_{i'}\right\}, \end{aligned}$$

where we have used the facts that

$$\begin{aligned}P(\alpha_i^{i*} \alpha_{i'}^{i*} = 0) &= 1, i \neq i'; \\P(\alpha_i^{i*} \alpha_{i'}^{i''} = 0) &= 1, i^* \neq i''; \\(\alpha_i^{i*})^2 &= \alpha_i^{i*}.\end{aligned}$$

Recalling now that

$$E(\alpha_i^{i*}) = \frac{1}{A}, \quad E(\alpha_i^{i*} \alpha_{i'}^{i''}) = \frac{1}{A(A-1)},$$

and $\sum_{i \neq i'} a_i a_{i'} = -\sum_i a_i^2$, we obtain

$$\begin{aligned}(\sum_i a_i^2) \left(\frac{1}{A} U - \frac{1}{UA} \sum_i u_i^2 + \frac{1}{UA(A-1)} \sum_{i \neq i'} u_i u_{i'} \right) \\= \frac{U}{(A-1)} (\sum_i a_i^2) \left[\frac{(A-1)}{A} - \frac{(A-1)}{A} U^* + \frac{1}{U^2 A} (U^2 - \sum_i u_i^2) \right] \\= U \sigma_a^2 \frac{1}{A} (A-1 - AU^* + U^* + 1 - U^*) \\= U(1 - U^*) \sigma_a^2.\end{aligned}$$

Hence the coefficient of σ_a^2 in the expectation of $A^* = 1/(a-1)A'$, the mean square for α , is

$$rUVW \frac{(1 - U^*)}{(a-1)}$$

as given in Table 1. Note that if all $u_{i*} = 1$, which would be the case if the total number of observations of each observed level i^* of α is the same, then $U = a$, $U^* = 1/a$, and the coefficient becomes rVW .

As another example we consider

$$\sum_{i*} u_{i*} E[p_{i*}^* \dots - p^* \dots]^2 = E \left\{ \sum_{i*} u_{i*} \left[\frac{1}{r u_{i*} V W} \sum_{j^* k^* f} \delta p_m - \frac{1}{r U V W} i^* \sum_{j^* k^* f} \delta p_m \right]^2 \right\},$$

where δ denotes $\delta_m^{i^* j^* k^* f}$

$$= \frac{1}{r^2 V^2 W^2} E \left\{ \sum_{i*} \frac{1}{u_{i*}} \left(\sum_{j^* k^* f} \delta p_m \right)^2 - \frac{1}{U} \left(\sum_{j^* k^* f} \delta p_m \right)^2 \right\}.$$

If the expressions in parentheses are expanded, it will be seen that a number of the terms will vanish because of relationships like

$$P(\delta_m^{i^* j^* k^* f} \delta_m^{i'' j'' k'' f'} = 0) = 1, \quad m \neq m';$$

and

$$P(\delta_m^{i^* j^* k^* f} \delta_m^{i'' j'' k'' f'} = 0) = 1, \quad (i^* j^* k^* f) \neq (i'' j'' k'' f').$$

If we also use the facts that

$$E(\delta_m^{i^*j^*k^*f}) = \frac{1}{P},$$

$$E(\delta_m^{i^*j^*k^*f} \delta_{m'}^{i'^*j'^*k'^*f'}) = \frac{1}{P(P-1)}, \quad m \neq m', (i^*j^*k^*f) \neq (i'^*j'^*k'^*f'),$$

$$\sum_m p_m^2 = -\sum_{m \neq m'} p_m p_{m'},$$

and if we use (j^*k^*f) to denote $\sum_{j^*k^*f^{(1)}}$, etc., then the expectation we seek to evaluate is

$$\begin{aligned} & \frac{\sum_m p_m^2}{r^2 V^2 W^2} \left\{ \sum_{i^*} \frac{1}{u_{i^*}} \left[\frac{1}{P} (j^*k^*f) - \frac{1}{P(P-1)} (j^*k^*f \neq f') \right. \right. \\ & - \frac{1}{P(P-1)} (j^* \neq j'^*k^*ff') - \frac{1}{P(P-1)} (j^*k^* \neq k'^*ff') \\ & - \left. \frac{1}{P(P-1)} (j^* \neq j'^*k^* \neq k'^*ff') \right] - \frac{1}{UP} \left[(i^*j^*k^*f) - \frac{1}{P-1} \right. \\ & \cdot \{ (i^*j^*k^*f \neq f') + (i^*j^* \neq j'^*k^*ff') \\ & + (i^*j^*k^* \neq k'^*ff') + (i^*j^* \neq j'^*k^* \neq k'^*ff') \\ & + (i^* \neq i'^*j^*k^*ff') + (i^* \neq i'^*j^* \neq j'^*k^*ff') \\ & \left. + (i^* \neq i'^*j^*k^* \neq k'^*ff') + (i^* \neq i'^*j^* \neq j'^*k^* \neq k'^*ff') \} \} \right\}. \end{aligned}$$

It remains only to write down the various values of the sums and collect terms. Thus

$$(j^*k^*f) = \sum_{j^*k^*f} (1) = \sum_{j^*k^*} n_{i^*j^*k^*} = ru_{i^*} \sum_{j^*k^*} v_{j^*} w_{k^*} = ru_{i^*} VW;$$

$$\begin{aligned} (j^*k^*f \neq f') &= \sum_{j^*k^*f \neq f'} (1) = \sum_{j^*k^*} n_{i^*j^*k^*} (n_{i^*j^*k^*} - 1) \\ &= r^2 u_{i^*}^2 \sum_{j^*} v_{j^*}^2 \sum_{k^*} w_{k^*}^2 - ru_{i^*} VW \\ &= r^2 u_{i^*}^2 V^2 W^2 - ru_{i^*} VW; \end{aligned}$$

$$\begin{aligned} (j^* \neq j'^*k^*ff') &= \sum_{j^*k^*f} \sum_{\substack{j'^*k'^*f' \\ j^* \neq j'^*}} (1) \\ &= \sum_{\substack{j^* \neq j'^* \\ k^*}} n_{i^*j^*k^*} n_{i^*j'^*k^*} = r^2 u_{i^*}^2 \sum_{j^* \neq j'^*} v_{j^*} v_{j'^*} \sum_{k^*} w_{k^*}^2 \\ &= r^2 u_{i^*}^2 (V^2 - V^2 V^*) W^2 W^*; \end{aligned}$$

$$(i^*j^*k^*f) = rUVW;$$

$$(i^*j^*k^*f \neq f') = r^2 U^2 U^* V^2 V^* W^2 W^* - rUVW;$$

$$(i^*j^* \neq j^*k^*ff') = r^2 U^2 U^* (V^2 - V^2 V^*) W^2 W^*;$$

etc. Thus the coefficient of $\sum_m p_m^2$ is

$$\begin{aligned} & \frac{1}{r^2 V^2 W^2} \frac{1}{P(P-1)} \left\{ \sum_{i^*} \frac{1}{u_{i^*}} [ru_{i^*} VW(P-1) - (r^2 u_{i^*}^2 V^* V^2 W^* W^2 - ru_{i^*} VW) \right. \\ & \quad - r^2 u_{i^*}^2 (V^2 - V^2 V^*) W^2 W^* - r^2 u_{i^*}^2 V^2 V^* (W^2 - W^2 W^*) \\ & \quad - r^2 u_{i^*}^2 V^2 W^2 (1 - V^*) (1 - W^*)] \\ & \quad - \frac{1}{U} [rUVW(P-1) - (r^2 U^2 U^* V^2 V^* W^2 W^* - rUVW) \\ & \quad - r^2 U^2 U^* V^2 (1 - V^*) W^2 W^* - r^2 U^2 U^* V^2 V^* W^2 (1 - W^*) \\ & \quad - r^2 U^2 U^* W^2 (1 - V^*) (1 - W^*) - r^2 U^2 (1 - U^*) V^2 V^* W^2 W^* \\ & \quad - r^2 U^2 (1 - U^*) V^2 (1 - V^*) W^2 W^* - r^2 U^2 (1 - U^*) V^2 V^* W^2 (1 - W^*) \\ & \quad \left. - r^2 U^2 V^2 W^2 (1 - U^*) (1 - V^*) (1 - W^*)] \right\} \\ & = \frac{1}{rVW} \frac{1}{P(P-1)} \left\{ \sum_{i^*} [(P-1) + 1 - ru_{i^*} VW \{ V^* W^* + (1 - V^*) W^* \right. \\ & \quad \left. + V^* (1 - W^*) + (1 - V^*) (1 - W^*) \}] \right. \\ & \quad - [(P-1) + 1 - rUVW \{ U^* V^* W^* + U^* (1 - V^*) W^* + U^* V^* (1 - W^*) \\ & \quad + U^* (1 - V^*) (1 - W^*) + (1 - U^*) V^* W^* + (1 - U^*) (1 - V^*) W^* \\ & \quad \left. + (1 - U^*) V^* (1 - W^*) + (1 - U^*) (1 - V^*) (1 - W^*) \}] \right\} \\ & = \frac{1}{rVW} \frac{1}{P(P-1)} (aP - P) = \frac{(a-1)}{rVW(P-1)}. \end{aligned}$$

Thus the coefficient of σ_p^2 in the expectation of A^* is

$$rVW \frac{(a-1)}{rVW} \frac{1}{(a-1)} = 1,$$

as given in Table 1.

In a similar way one can complete the evaluation of $E(A')$, proceeding from component to component; and of course the other mean squares may be handled in the same fashion. In view of the symmetry of factors one can write down, at once, $E(B^*)$ and $E(C^*)$ from the results for $E(A^*)$, and likewise for I_{AB}^* , I_{AC}^* , and I_{BC}^* . A check on results is that the expectation of the total sum of squares should equal the sum of the expectations.

The complexity of the formulae and also of the algebra is considerably simplified when the number of observations per cell is a constant, say r .

While the operations with the statistical model may appear tedious, this is more apparent than real, in that with some familiarity with the technique a good deal of the writing can be decreased through short-cut notation and simplifica-

tion by inspection. Furthermore the operations are quite elementary and mechanical; and in addition to the symmetries we have mentioned, there are others, such as the symmetry with respect to σ_{ab}^2 and σ_{ac}^2 in $E(A^*)$.

13. A more symmetric form for ems; extension of results. The general formulae for ems may be put in a more symmetric form which is simpler in appearance and which makes very simple the extension of the results to four or more factors. The modified form of the results involves certain linear combination of the defined components of variation in terms of which the expectations of mean squares have the *appearance* corresponding to an "all factors random, number of units infinite" situation. This general pattern for ems, involving appropriate and definite rules for forming the linear combinations of components of variation, has been obtained by one or both of the present authors for more complex designs and situations than we have studied in this paper; ramifications will be discussed in later communications.

We shall consider, for definiteness, the results of Table 1 on ems for the case of proportional numbers. These results are given in Table 5 in terms of the following notation:

$$\Sigma_a = \sigma_a^2 - \frac{1}{B} \sigma_{ab}^2 - \frac{1}{C} \sigma_{ac}^2 - \frac{1}{P} Q_{ap}^2 + \frac{1}{BC} \sigma_{abc}^2 + \frac{1}{BP} Q_{abp}^2 + \frac{1}{CP} Q_{acp}^2 - \frac{1}{BCP} Q_{abc p}^2.$$

Σ_b and Σ_c are defined analogously.

$$\Sigma_{ab} = \sigma_{ab}^2 - \frac{1}{C} \sigma_{abc}^2 - \frac{1}{P} Q_{abp}^2 + \frac{1}{CP} Q_{abc p}^2.$$

Σ_{ac} and Σ_{bc} are defined analogously.

$$\Sigma_{abc} = \sigma_{abc}^2 - \frac{1}{P} Q_{abc p}^2.$$

$$\begin{aligned} \Sigma_p = \sigma_p^2 - \frac{1}{A} Q_{ap}^2 - \frac{1}{B} Q_{bp}^2 - \frac{1}{C} Q_{cp}^2 + \frac{1}{AB} Q_{abp}^2 + \frac{1}{AC} Q_{acp}^2 \\ + \frac{1}{BC} Q_{bcp}^2 - \frac{1}{ABC} Q_{abc p}^2. \end{aligned}$$

$$\Sigma_{ap} = Q_{ap}^2 - \frac{1}{B} Q_{abp}^2 - \frac{1}{C} Q_{acp}^2 + \frac{1}{BC} Q_{abc p}^2.$$

Σ_{bp} and Σ_{cp} are defined analogously.

$$\Sigma_{abp} = Q_{abp}^2 - \frac{1}{C} Q_{abc p}^2.$$

Σ_{acp} and Σ_{bcp} are defined analogously.

$$\Sigma_{abc p} = Q_{abc p}^2.$$

$$\begin{aligned} \Sigma_0 = \sigma^2 + \Sigma_{abc p} + \Sigma_{bcp} + \Sigma_{acp} + \Sigma_{abp} + \Sigma_{cp} + \Sigma_{bp} + \Sigma_{ap} + \Sigma_p \\ = \sigma^2 + \sigma_p^2 + \sigma_q^2. \end{aligned}$$

TABLE 5
Symmetric form for the results of Table 1

Mean squares	Expected mean squares
A^*	$rUVW \frac{(1 - U^*)}{(a - 1)} (\Sigma_a + V^*\Sigma_{ab} + W^*\Sigma_{ac} + V^*W^*\Sigma_{abc}) + \Sigma_0$
B^*	$rUVW \frac{(1 - V^*)}{(b - 1)} (\Sigma_b + U^*\Sigma_{ab} + W^*\Sigma_{bc} + U^*W^*\Sigma_{abc}) + \Sigma_0$
C^*	$rUVW \frac{(1 - W^*)}{(c - 1)} (\Sigma_c + U^*\Sigma_{ac} + V^*\Sigma_{bc} + U^*V^*\Sigma_{abc}) + \Sigma_0$
I_{AB}^*	$rUVW \frac{(1 - U^*)(1 - V^*)}{(a - 1)(b - 1)} (\Sigma_{ab} + W^*\Sigma_{abc}) + \Sigma_0$
I_{AC}^*	$rUVW \frac{(1 - U^*)(1 - W^*)}{(a - 1)(c - 1)} (\Sigma_{ac} + V^*\Sigma_{abc}) + \Sigma_0$
I_{BC}^*	$rUVW \frac{(1 - V^*)(1 - W^*)}{(b - 1)(c - 1)} (\Sigma_{bc} + U^*\Sigma_{abc}) + \Sigma_0$
I_{ABC}^*	$rUVW \frac{(1 - U^*)(1 - V^*)(1 - W^*)}{(a - 1)(b - 1)(c - 1)} \Sigma_{abc} + \Sigma_0$
R^*	Σ_0

An inverse relationship giving the σ^2 and Q^2 quantities explicitly in terms of the Σ 's is easily written down.

The form of the results given in Table 5 not only makes entirely clear the pattern for extension to more than three factors but also indicates what are, in general, the estimable quantities in the analysis of variance. It will be evident that an unbiased estimate, based on the analysis of variance mean squares, always exists for each Σ quantity in Table 5. It is of interest that the Σ quantities depend only on the population sizes and not on the sample sizes.

To make explicit the pattern of extensions to more than three factors, we give $E(I_{AB}^*)$ when we have four factors \mathfrak{A} , \mathfrak{B} , \mathfrak{C} , \mathfrak{D} . The notation and definitions implicit should be clear. We use X as analogous to U , V , W , and X^* as analogous to U^* , V^* , W^* , with definitions of components of variation as before. Then

$$E(I_{AB}^*) = \Sigma_0 + rUVWX \frac{(1 - U^*)(1 - V^*)}{(a - 1)(c - 1)} \cdot (W^*X^*\Sigma_{abcd} + W^*\Sigma_{abc} + X^*\Sigma_{abd} + \Sigma_{ab}),$$

where

$$\Sigma_{abcd} = Q_{abcd}^2 - \frac{1}{P} Q_{abcdp}^2,$$

$$\Sigma_{abc} = \sigma_{abc}^2 - \frac{1}{D} \sigma_{abcd}^2 - \frac{1}{P} Q_{abcp}^2 + \frac{1}{DP} Q_{abcdp}^2,$$

$$\begin{aligned} \Sigma_{ab} = \sigma_{ab}^2 - \frac{1}{C} \sigma_{abc}^2 - \frac{1}{D} \sigma_{abd}^2 - \frac{1}{P} Q_{abp}^2 + \frac{1}{CD} \sigma_{abcd}^2 \\ + \frac{1}{CP} Q_{abcp}^2 + \frac{1}{DP} Q_{abdp}^2 - \frac{1}{CDP} Q_{abcdp}^2, \end{aligned}$$

$$\Sigma_0 = \Sigma_p + \Sigma_{ap} + \Sigma_{bp} + \dots + \Sigma_{abcdp} + \sigma^2,$$

etc.

Further discussion on the extension of results for expected mean squares to other designs, on a more formal (operational) statement of definition of the Σ quantities, and on the formal general reciprocal definition of the σ^2 and Q^2 quantities in terms of the Σ 's is deferred to a later publication.

14. Estimation of effects, interactions and errors. In many factorial experiments one of the objectives of the experiment will be the estimation of contrasts, such as $\sum_i k_i a_i$, with $\sum_i k_i = 0$, and in particular differences such as $a'_i - a'_i$; also the uncertainty (usually as measured by variance) of such estimates needs to be estimated. The essential objective of this section is to illustrate briefly the use of the statistical model in such "linear estimation" problems.

To simplify the exposition we shall deal with the case of two factors, which is equivalent, formally, to putting $C = c = 1$ for the situation we developed earlier. We can now drop the subscripts k and k^* and all interactions involving c . The population model becomes

$$y_{ijm} = \mu + a_i + b_j + (ab)_{ij} + p_m + q_{ijm} + \epsilon_{ijm}.$$

The statistical model becomes

$$\begin{aligned} x_{i^*j^*f} = \mu + \sum_i \alpha_i^* a_i + \sum_j \beta_j^* b_j + \sum_{ij} \alpha_i^* \beta_j^* (ab)_{ij} + \sum_m \delta_m^{i^*j^*f} p_m \\ + \sum_{ijm} \alpha_i^* \beta_j^* \delta^{i^*j^*f} (q_{ijm} + \epsilon_{ijm}). \end{aligned}$$

We recall that our experiment involved the random selection of a levels from A of factor \mathfrak{A} and b levels from B of \mathfrak{B} , where $a \leq A$, $b \leq B$, and the random allocation of the ab selected treatment combinations to randomly selected experimental units (from a population of size P), so that each selected treatment (i^*j^*) appeared $n_{i^*j^*}$ times, $n_{i^*j^*} \geq 1$.

For the case of $A > a$, $B > b$ the association of $n_{i^*j^*}$ values with population treatment combinations (ij) is a random one. For the case of \mathfrak{A} and \mathfrak{B} fixed factors one of two situations might exist, namely when i^* and i (and j^* and j) are taken as the same index or when the range of $i^*(j^*)$ is a random permutation of the range of $i(j)$. In the first case we can speak of having n_{ij} observations for treatment combination ij ; in the second case we have $(\sum_{i^*j^*} \alpha_i^* \beta_j^* n_{i^*j^*})$ observations, a random variable having average value $1/ab \sum_{i^*j^*} n_{i^*j^*}$, for treatment (ij). To bypass this difficulty, we shall consider in this paper the case of equal numbers, i.e., $n_{i^*j^*} = r \geq 1$, all i^*j^* .

Under this last condition,

$$x_{i^{\cdot\cdot}} = \mu + a_{i^{\cdot}}^* + b^* + (ab)_{i^{\cdot}}^* + p_{i^{\cdot}}^* + q_{i^{\cdot}}^* + \epsilon_{i^{\cdot}}^*$$

where

$$a_{i^{\cdot}}^* = \sum_i \alpha_i^* a_i, \quad b^* = \frac{1}{b} \sum_{j^*} \beta_j^* b_j;$$

$$(ab)_{i^{\cdot}}^* = \frac{1}{b} \sum_{j^*} \alpha_i^* \beta_j^* (ab)_{ij}; \quad p_{i^{\cdot}}^* = \frac{1}{rb} \sum_{j^*f} \delta_m^{i^*j^*f} p_m;$$

etc.

With no further restrictions an $a \leq A$, $b \leq B$, let us put

$$x^{i^{\cdot\cdot}} = \frac{\sum_{i^*} \alpha_i^{i^*} x_{i^{\cdot\cdot}}}{\sum_{i^*} \alpha_i^{i^*}}$$

when the right-hand side is *determinate*. This quantity will be *indeterminate* whenever population level i of α is *not* included among those actually selected, for then both numerator and denominator above will be zero. Then, when $x^{i^{\cdot\cdot}}$ exists, the denominator above is 1 and

$$\begin{aligned} x^{i^{\cdot\cdot}} &= \mu + a^i + \sum_{i^*} \alpha_i^{i^*} [b^* + (ab)_{i^{\cdot}}^* + p_{i^{\cdot}}^* + q_{i^{\cdot}}^* + \epsilon_{i^{\cdot}}^*] \\ &= \mu + a_i + \frac{1}{b} \sum_{j^*} \beta_j^* b_j + \frac{1}{b} \sum_{j^*} \beta_j^* (ab)_{ij} + \sum_{i^*} \alpha_i^{i^*} [p_{i^{\cdot}}^* + q_{i^{\cdot}}^* + \epsilon_{i^{\cdot}}^*]. \end{aligned}$$

It should be noted that this statistical model for $x^{i^{\cdot\cdot}}$ is *conditional* on level i of α , having been one of the selected a levels of α ; hence, in this expression, we take $P(\alpha_i^* = 1) = 1/a$, which is the conditional probability that selected level i^* corresponds to population level i , given that i is selected.

In the last expression, all terms after the first two on the right-hand side have expectation zero, whatever the relation of B to b . For example

$$\begin{aligned} E[\sum_{j^*} \beta_j^* (ab)_{ij}] &= \frac{1}{B} \sum_{j^*} [\sum_j (ab)_{ij}] = 0; \\ E[\sum_{i^*} \alpha_i^{i^*} p_{i^{\cdot}}^*] &= E\left[\frac{1}{rb} \sum_{i^*} \alpha_i^{i^*} \sum_{j^*f} \delta_m^{i^*j^*f} p_m\right] = \frac{1}{rbaP} \sum_{i^*f} (\sum_m p_m) = 0; \\ E[\sum_{i^*} \alpha_i^{i^*} q_{i^{\cdot}}^*] &= E\left[\frac{1}{rb} \sum_{i^*} \alpha_i^{i^*} \sum_{j^*f} \sum_{jm} \alpha_i^{i^*} \beta_j^* \delta_m^{i^*j^*f} q_{ijm}\right] \\ &= E\left[\frac{1}{rb} \sum_{i^*j^*f} \sum_{jm} \alpha_i^{i^*} \beta_j^* \delta_m^{i^*j^*f} q_{ijm}\right] \\ &= \frac{1}{rb} \frac{1}{a} \frac{1}{B} \frac{1}{P} \sum_{i^*j^*f} \sum_j (\sum_m q_{ijm}) = 0; \end{aligned}$$

etc.

Thus $x^{i\cdots}$ is an unbiased estimate of

$$\mu + a_i = Y_{i\cdots},$$

which is the conceptual over-all mean "true" response from all (population) treatment combinations involving level i of \mathcal{A} on all (population) experimental units. Hence, an unbiased estimate of the difference of the main effects of levels i and i' of \mathcal{A} , $a_i - a_{i'}$, is given by $(x^{i\cdots} - x^{i'\cdots})$, when both of these quantities are determinate, independent of whether either \mathcal{A} or \mathcal{B} is fixed or not, and independent of whether interactions of factors with each other or with units are negligible or not.

It may be appropriate here to emphasize that the difference $(a_i - a_{i'})$ is independent of the other levels of \mathcal{A} under study but is *very much dependent*, in general, on what population of levels of \mathcal{B} and of experimental units is under consideration. (Note that the preceding sentence refers to population parameters and *not* to sample estimates.)

In considering uncertainties involved in the estimation of $(a_i - a_{i'})$ by $(x^{i\cdots} - x^{i'\cdots})$ it is clear from the model for $x^{i\cdots}$ that the estimate will be affected by the interactions of levels i and i' of \mathcal{A} with levels of \mathcal{B} only if \mathcal{B} is not fixed, for if \mathcal{B} is a fixed factor, then $\sum_{j^*j} \beta_j^*(ab)_{ij} = \sum_i (ab)_{ij} = 0$, independent of i . On the other hand if \mathcal{B} is *not* a fixed factor, then the term

$$\frac{1}{b} \sum_{j^*j} \beta_j^* [(ab)_{ij} - (ab)_{i'j}]$$

does not vanish from $(x^{i\cdots} - x^{i'\cdots})$.

If factor \mathcal{B} is fixed and, further, unit treatment interactions are negligible, i.e., all $q_{ijm} = 0$, then

$$x^{i\cdots} - x^{i'\cdots} = a_i - a_{i'} + \sum_{i^*} (\alpha_i^* - \alpha_{i'}^*) (p_{i^*} + \epsilon_{i^*}^*).$$

The variance of this estimate is

$$\begin{aligned} E[\sum_{i^*} (\alpha_i^* - \alpha_{i'}^*) (p_{i^*} + \epsilon_{i^*}^*)]^2 &= \frac{2\sigma^2}{rb} + E[\sum_{i^*} (\alpha_i^* - \alpha_{i'}^*) \left(\frac{1}{rb} \sum_{j^*j} \delta_m^{i^*j^*f} p_m \right)]^2 \\ &= \frac{2\sigma^2}{rb} + \frac{1}{r^2b^2} E \left[\left(\sum_{i^*j^*f} \alpha_i^* \delta_m^{i^*j^*f} p_m^2 + \sum_{i^*j^*f \neq i'^*j'^*f'} \alpha_i^* \delta_m^{i^*j^*f} \delta_m^{i'^*j'^*f'} p_m p_{m'} \right. \right. \\ &\quad \left. \left. + \sum_{i^*j^*f \neq i'^*j'^*f'} \alpha_i^* \delta_m^{i^*j^*f} \delta_m^{i'^*j'^*f'} p_m p_{m'} \right) + (\text{similar terms with } i' \text{ for } i) \right. \\ &\quad \left. - 2 \left(\sum_{i^*j^*f} \sum_{i'^*j'^*f'} \alpha_i^* \alpha_{i'}^* \delta_m^{i^*j^*f} \delta_m^{i'^*j'^*f'} p_m p_{m'} \right. \right. \\ &\quad \left. \left. + \sum_{i^*j^*f \neq i'^*j'^*f'} \sum_{i'^*j'^*f' \neq i^*j^*f} \alpha_i^* \alpha_{i'}^* \delta_m^{i^*j^*f} \delta_m^{i'^*j'^*f'} p_m p_{m'} \right) \right] \\ &= \frac{2\sigma^2}{rb} + \frac{2}{r^2b^2} \left[\left(\frac{rb}{P} - \frac{r(r-1)b}{P(P-1)} - \frac{r^2b(b-1)}{P(P-1)} \right) \right] \end{aligned}$$

$$+ \left(\frac{r^2 b}{P(P-1)} + \frac{r^2 b(b-1)}{P(P-1)} \right) \sum_m p_m^2$$

$$= \frac{2}{rb} (\sigma^2 + \sigma_p^2).$$

Hence under the conditions that (i) unit-treatment interactions are zero and (ii) B is a fixed factor (i.e., $B = b$), the variance of the estimate of the difference of the main effects of two levels of α is estimated unbiasedly by $2R^*/rb$, where R^* is the "residual mean square" in the analysis of variance.

We consider next the variance of the estimate ($x^{i''} - x^{i'j''}$) without the above restrictions. Then

$$\text{var } (x^{i''} - x^{i'j''}) = \frac{2}{rb} (\sigma^2 + \sigma_p^2)$$

$$+ E \left[\frac{1}{b} \sum_{j,j'} \beta_j^{j''} ((ab)_{ij} - (ab)_{i'j'}) + \sum_{i''} (\alpha_{i''} - \alpha_{i'j''}) q_{i''}^{*..} \right]^2$$

$$= \frac{2}{rb} (\sigma^2 + \sigma_p^2) + \frac{1}{b^2} E \left\{ \sum_{j,j'} \beta_j^{j''} [(ab)_{ij} - (ab)_{i'j'}]^2 \right.$$

$$+ \sum_{\substack{j \neq j'' \\ j' \neq j''}} \beta_j^{j''} \beta_{j'}^{j''} [(ab)_{ij} - (ab)_{i'j'}][(ab)_{ij'} - (ab)_{i'j'j''] \}$$

$$+ \frac{1}{r^2 b^2} E \left[\sum_{i''} \alpha_{i''}^* \sum_{j,m} \beta_j^{j''} \delta_m^{i''j''} q_{ijm} - \sum_{i''} \alpha_{i''}^* \sum_{j',m} \beta_{j'}^{j''} \delta_m^{i''j''} q_{i'j'm} \right]^2$$

$$= \frac{2}{rb} (\sigma^2 + \sigma_p^2) + \frac{1}{b^2} \left\{ \frac{b}{B} \sum_j [(ab)_{ij}^2 + (ab)_{i'j}^2 - 2(ab)_{ij}(ab)_{i'j}] \right.$$

$$- \frac{b(b-1)}{B(B-1)} \sum_j [(ab)_{ij}^2 + (ab)_{i'j}^2 - 2(ab)_{ij}(ab)_{i'j}] \}$$

$$+ \frac{1}{r^2 b^2} \left\{ \left[\frac{rb}{BP} - \frac{r(r-1)b}{P(P-1)B} \right] \sum_{j,m} q_{ijm}^2 - \frac{r^2 b(b-1)}{P(P-1)B(B-1)} \sum_{j \neq j'} q_{ijm} q_{i'j'm} \right.$$

$$+ \left[\frac{rb}{PB} - \frac{r(r-1)b}{P(P-1)B} \right] \sum_{j,m} q_{i'jm}^2 - \frac{r^2 b(b-1)}{P(P-1)B(B-1)} \sum_{j \neq j'} q_{i'jm} q_{i'j'm} \}$$

$$+ \frac{2r^2 b}{P(P-1)B} \sum_{j,m} q_{ijm} q_{i'jm} + \frac{2r^2 b(b-1)}{P(P-1)B(B-1)} \sum_{j \neq j'} q_{ijm} q_{i'j'm} \}$$

$$= \frac{2}{rb} (\sigma^2 + \sigma_p^2) + \frac{1}{b} \frac{(B-b)}{B(B-1)} \sum_j [(ab)_{ij} - (ab)_{i'j}]^2$$

$$+ \frac{1}{rb} \frac{1}{BP(P-1)} [(P-r) \sum_{j,m} (q_{ijm}^2 + q_{i'jm}^2) + 2r \sum_{j,m} q_{ijm} q_{i'jm}]$$

$$\begin{aligned}
 & - \frac{(b-1)}{bP(P-1)B(B-1)} \sum_{j \neq j'} \sum_m (q_{ijm} q_{ij'm} + q_{i'jm} q_{i'j'm} - 2q_{ijm} q_{i'j'm}) \\
 = & \frac{2}{rb} (\sigma^2 + \sigma_p^2) + \frac{(B-b)}{bB} \frac{1}{(B-1)} \sum_j [(ab)_{ij} - (ab)_{i'j}]^2 \\
 & + \frac{1}{rb} \frac{1}{B(P-1)} \sum_{jm} (q_{ijm}^2 + q_{i'jm}^2) - \frac{1}{b} \frac{1}{PB(P-1)} \sum_{jm} (q_{ijm} - q_{i'jm})^2 \\
 & - \frac{(b-1)}{bP(P-1)B(B-1)} \sum_{j \neq j'} \sum_m (q_{ijm} - q_{i'jm})(q_{ij'm} - q_{i'j'm}).
 \end{aligned}$$

Before considering the estimation of this variance we shall obtain a useful related quantity, namely, the average variance of estimates such as $(x^{i\cdots} - x^{i'\cdots})$

$$\begin{aligned}
 \frac{1}{A(A-1)} \sum_{i \neq i'} \text{Var} (x^{i\cdots} - x^{i'\cdots}) \\
 = & \frac{2}{rb} (\sigma^2 + \sigma_p^2 + \sigma_a^2) + \frac{2}{b} \left(\frac{B-b}{B} \left[\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2 \right] - \frac{2}{P} Q_{ap}^2 \right) \\
 = & \frac{2}{rb} [E(A^*) - rb\sigma_a^2].
 \end{aligned}$$

This displays explicitly (what is after all obvious) the relationship of the average variance of estimates of differences and the analysis of variance ems. Clearly this average variance can be estimated unbiasedly only if conditions are such that an "unbiased error term" exists for σ_a^2 , which will be true only when unit treatment interactions are negligible or/and the population of experimental units is very large. In general, the estimate of this average variance based on the error term for σ_a^2 given in Table 2 will be positively biased, i.e., the variance will tend to be overestimated. With regard to the component σ_{ab}^2 we note that its importance in this formula for the average variance of estimates of differences of main effects of levels of \mathcal{B} is determined by the relationship of B (the population size of levels of factor \mathcal{B} to b (the sample size for levels of \mathcal{B}) and *not* by any considerations concerning A and a .

If the average variance given above was felt not to be adequate as an estimate of the variance of a specific difference $(x^{i\cdots} - x^{i'\cdots})$ one could carry out what would amount to an analysis of variance involving only the observations of relevance, namely, those that go into $(x^{i\cdots} - x^{i'\cdots})$. Thus if we extend our previous notation to

$$x^{ij*} = \frac{\sum_{i^*} \alpha_i^{i^*} x_{i^*j^*}^{i^*j^*}}{\sum_{i^*} \alpha_i^{i^*}}, \quad \text{and} \quad x^{ij*} = \frac{1}{r} \sum_i x^{ij*},$$

then the sums of squares of this "partial" analysis of variance would be

$$\frac{2}{rb} (x^{i\cdots} - x^{i'\cdots})^2; \quad \frac{r}{2} \sum_{j^*} (x^{ij*} + x^{i'j*} - x^{i\cdots} - x^{i'\cdots})^2;$$

$$\frac{r}{2} \sum_{j''} (x^{ij''} - x^{i'j''} - x^{i''} + x^{i'j''})^2; \sum_{j''f} [(x^{ij''f} - x^{i'j''f})^2 + (x^{i'j''f} - x^{i'j''})^2];$$

i.e., sums of squares for level i versus level i' of \mathcal{A} ; sum of squares for levels of \mathcal{B} averaged over levels i and i' of \mathcal{A} ; sums of squares for interactions of levels of \mathcal{B} with levels i and i' of \mathcal{A} ; and residual. Clearly this partial analysis of variance will bear the same relation to the variance of $(x^{i''} - x^{i'j''})$ as the complete analysis bears to the average variance of differences.

When interactions with experimental units are negligible, the residual mean square from the partial analysis of variance will have the same expectation as that for the complete one.

When $B > b$ and the interactions of levels i and i' of \mathcal{A} with levels of \mathcal{B} may be considerably different from the interactions of other levels of \mathcal{A} with levels of \mathcal{B} , it may be worth while carrying out the partial analysis of variance to obtain estimates of the variance of the specific difference.

The preceding discussion can be applied symmetrically to factor \mathcal{B} and be extended to a three or more factor situation. Similarly the statistical model can be employed formally to answer questions involving the estimation of specific interactions, or differences of such, and to find variances of such estimates.

So far as experimental unit variability is concerned, randomization is fully effective in providing unbiased linear estimates and in giving unbiased estimates of the component of variation corresponding to the additive unit errors; but, in general, randomization does not lead to the unbiased estimation of the contribution to variances of linear estimates due to the interactive unit error. It is, however, probably true that in many situations this latter bias will not be important.

REFERENCES

- [1] J. CORNFIELD AND J. W. TUKEY, "Average values of mean squares in factorials," *Ann. Math. Stat.*, Vol. 27, (1956), pp. 907-949.
- [2] D. R. COX, Some Statistical Methods Connected with Series of Events. *J. Roy. Stat. Soc. B*, 17, 129-157, 1955.
- [3] R. A. FISHER, *Statistical methods for research workers*, Oliver and Boyd, Edinburgh, 1st Ed., 1925.
- [4] R. A. FISHER, "The arrangement of field experiments," *J. Min. Agric.*, Vol. 33 (1926), pp. 503-513.
- [5] R. A. FISHER, "Discussion of 'Complex Experiments' by F. Yates," *J. Roy. Stat. Soc.*, Suppl., Vol. 2 (1935), pp. 229-231.
- [6] R. A. FISHER, *The design of experiments*, Oliver and Boyd, Edinburgh, 1st ed., 1935.
- [7] R. HOOKE, "Symmetric Functions of a Two-way Array," *Annals Math. Stat.*, Vol. 27, (1956), pp. 55-79.
- [8] H. SCHEFFÉ, "Alternate models for the analysis of variance," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 251-271.
- [9] H. F. SMITH, "Analysis of variance with unequal but proportionate numbers of observations in the sub-classes of a two-way classification," *Biometrics*, Vol. 7 (1951), pp. 70-74.
- [10] H. F. SMITH, *Variance components, finite populations and statistical inference*, North Carolina Institute of Statistics Mimeo Series 135 (1955).

- [11] J. W. TUKEY, Variances of Variance Components. I. Balanced Designs. II. The Unbalanced Single Classification. III. Third moments in a Balanced Single Classification. These papers are to appear in *Annals Math. Stat.*
- [12] M. B. WILK, "The randomization analysis of a generalized randomized block design," M.S. Thesis, Iowa State College, 1953.
- [13] M. B. WILK, "The randomization analysis of a generalized randomized block design," *Biometrika*, Vol. 42 (1955), pp. 70-79.
- [14] M. B. WILK, "Linear models and randomized experiments," Ph.D. Thesis, Iowa State College, 1955.
- [15] M. B. WILK AND O. KEMPTHORNE, "Fixed, mixed and random models," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1144-1178.
- [16] F. YATES, "The principles of orthogonality and confounding in replicated experiments," *J. Agric. Sci.*, 1932.
- [17] F. YATES, "Complex experiments," *J. Roy. Stat. Soc., Suppl.*, Vol. 2 (1935), pp. 181-247.
- [18] F. YATES, *The design analysis of factorial experiments*, TC 35, Imperial Bureau of Soil Science, Harpend, 1937.

ON A MEASURE OF THE INFORMATION PROVIDED BY AN EXPERIMENT^{1, 2}

By D. V. LINDLEY

University of Cambridge and University of Chicago

1. Summary. A measure is introduced of the information provided by an experiment. The measure is derived from the work of Shannon [10] and involves the knowledge prior to performing the experiment, expressed through a prior probability distribution over the parameter space. The measure is used to compare some pairs of experiments without reference to prior distributions; this method of comparison is contrasted with the methods discussed by Blackwell. Finally, the measure is applied to provide a solution to some problems of experimental design, where the object of experimentation is not to reach decisions but rather to gain knowledge about the world.

2. Introduction. Shannon has introduced two important ideas into the theory of information in communications engineering. The first idea is that information is a statistical concept. The statistical frequency distribution of the symbols that make up a message must be considered before the notion can be discussed adequately. The second idea springs from the first and implies that on the basis of the frequency distribution, there is an essentially unique function of the distribution which measures the amount of the information. It is the purpose of the present paper to apply these two ideas to statistical theory by discussing the notion of information in an experiment, rather than in a message. The second of Shannon's ideas has been applied to statistical theory by Kullback and Leibler [6], [7], [8]; but our application is quite distinct from theirs. The interpretation of Shannon's ideas in current statistical theory has been given by McMillan [9]. The discussion in that paper is related to, and partly inspired, that given here. A referee has kindly pointed out that Shannon's theory has been applied in psychometric problems by L. J. Cronbach in an unpublished report [14]. Definition 2, in particular, is used by Cronbach.

The situation in communications engineering is that there is a transmitted message, x , which is received as a message, y . By considerations of the informations in x and y it is possible to discuss the rate at which information has been transmitted along the channel. The analogous description in statistical theory is provided by replacing x by the knowledge of the state of nature, usually expressed by the knowledge of a finite number of parameters, prior to an experiment, and by replacing y by the knowledge after the experiment. The com-

Received August 2, 1955.

¹ Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Office of Naval Research.

² This is a revised version of papers presented at the Chapel Hill and Berkeley meetings of the Institute of Mathematical Statistics in April and July, 1955.

parison of the knowledge before and after the experiment makes it possible to discuss the amount of information provided by the experiment. The average of this, for fixed prior knowledge, determines the average amount of information. The measure of information is given by Shannon's function. But, just as it is essential to consider the statistical character of the message x , so it is necessary to consider the statistical character of the knowledge of the state of nature. Prior probability distributions are therefore basic to the study. It seems obvious to the author that prior distributions, though usually anathema to the statistician, are essential to the notion of experimental information. To take an extreme case, if the prior distribution is concentrated on a single parameter value, that is, if the state of nature is known, then no experiment can be informative.

It may happen that, whatever the prior knowledge, one experiment is more informative than another. We shall meet such examples below. In this case it is possible to compare the two experiments absolutely, without reference to prior knowledge. Methods of comparing experiments have been suggested by Bohnenblust, Shapley, and Sherman (described by Blackwell in [2]) and by Blackwell [2]. These methods of comparison are contrasted with the one presented here, and it is shown that if one experiment is more informative than another by Blackwell's criterion, then it is also true of that used here; the converse is false.

The Bohnenblust method of comparison is formulated in decision theory language and involves considerations of losses. These notions are not used here; the concepts used are perhaps more related to the inference problem than to the decision problem (see Barnard [1]). In this paper it is suggested that although indisputably one purpose of experimentation is to reach decisions, another purpose is to gain knowledge about the state of nature (that is, about the parameters) without having specific actions in mind. The knowledge is measured by the amount of information, as described above. The following rule of experimentation is therefore suggested: perform that experiment for which the expected gain in information is the greatest, and continue experimentation until a preassigned amount of information has been attained. The consequences of this rule are explored and shown, for example, to lead to sequential probability ratio tests. Binomial and normal sampling are also considered as special cases.

3. The experiment will result in an observation, x , belonging to a space, \mathbf{X} . The space \mathbf{X} has a σ -field, \mathfrak{B} , of subsets, X . For every θ belonging to a space Θ is defined a probability measure on \mathfrak{B} . We shall suppose that as θ ranges through Θ the probability measures on \mathfrak{B} are all absolutely continuous with respect to a fixed measure on \mathfrak{B} . This permits us to describe each probability measure by a probability density function $p(x | \theta)$, such that the probability measure of a subset X is given by $\int_X p(x | \theta) dx$, where, for simplicity of notation, we have denoted integration with respect to the dominating measure by

dx . The ordered quadruple³ $\varepsilon = \{X, \mathcal{G}, \Theta, P\}$, where P is the set of $p(x | \theta)$, characterizes an experiment, ε . Again, for simplicity in notation, we shall not distinguish between random variables and the values assumed by them, nor shall we attempt to be specific in describing the density functions. Thus, $p(x)$ will denote the density function of the random variable x ; similarly, $p(\theta)$ will denote the density function of θ , without any suggestion that the random variables x and θ have the same density. These devices avoid such clumsy notation as $p_x(y)$ for the density of the random variable x when x assumes the numerical value y .

We shall suppose that Θ is endowed with a σ -field of subsets; usually, Θ will be a subset of n -dimensional Euclidean space and the σ -field will be the Borel field. A prior distribution for θ will be a probability measure on this field, and again we shall suppose it to be described by a probability density function $p(\theta)$ with respect to a measure denoted by $d\theta$. Thus, in accord with the notational conventions described above, we have, for example,

$$(1) \quad p(x) = \int_{\Theta} p(x | \theta) p(\theta) d\theta,$$

and Bayes' theorem reads

$$(2) \quad p(\theta | x) = p(x | \theta) p(\theta) / p(x).$$

The ranges of integration in the following formulas will always be the whole space, either X or Θ , and will be omitted.

For a prior distribution $p(\theta)$, the amount of information with respect to $d\theta$ is defined to be

$$(3) \quad s_0 = \int p(\theta) \log p(\theta) d\theta$$

whenever the integral exists. For any θ for which $p(\theta) = 0$, define $p(\theta) \log p(\theta)$ to be zero. A useful alternative notation is

$$(4) \quad s_0 = E_0 \log p(\theta),$$

where E_0 denotes the expectation operator with respect to θ .

The reasons for the introduction of this function have been given by Shannon. Translated into the language of experimentation, the basic reason is this: Consider the case where Θ is finite; then the amount of information, I , in a prior distribution can be measured by how much information it is necessary to provide before the value of θ is known. This latter information could be provided in two stages. For the first, let Θ_1 be a non-empty proper subset of Θ with $P = \int_{\Theta_1} p(\theta) d\theta \neq 0$ or 1, and suppose the experimenter is told whether $\theta \in \Theta_1$ or its complement. This provides amount I_1 , say; the prior distribution being $(P, 1 - P)$. In the second stage, suppose the experimenter is told the value of

³ Strictly, the quadruple should be a quintuple and should include the dominating measure; for convenience, it will be omitted.

θ ; the information provided is I_2 or I_3 , say, according as he knew $\theta \in \Theta_1$ or its complement. (The necessary distributions are $p(\theta)/P$ and $p(\theta)/(1 - P)$, respectively.) Then Shannon requires that the information provided in the first stage and the average amount provided in the second stage add up to the total information; that is,

$$I = I_1 + PI_2 + (1 - P)I_3.$$

This additivity requirement is the fundamental postulate. It finds its general form in Theorem 2, below. Shannon then shows ([10], Appendix 2) that $I = \sum p(\theta) \log p(\theta)$, apart from an arbitrary multiplying constant, is the only function having this property together with a mild continuity property.

We note that the amount of information, so defined, is not invariant under a change of description of the parameter space. This lack of invariance need cause no concern, as it will disappear when the expression is used to define the average information in the experiment. The minus sign introduced by Shannon in front of the integral is not used. The reason for this is as follows: the maximum information, in a statistician's sense, will be obtained when the probability distribution is concentrated on a single value of θ , and the information will be reduced as the distribution of θ "spreads"; this is exactly the reverse of the situation faced by a communications engineer, where the concentration on a single value would allow no choice in his messages. The two scales are therefore reversed.

After the experiment has been performed and the value x observed, the posterior distribution of θ is $p(\theta | x)$, given by (2), and the amount of information is

$$(5) \quad s_1(x) = \int p(\theta | x) \log p(\theta | x) d\theta.$$

(If $p(\theta | x) = 0$, define the integrand to be zero.)

DEFINITION 1. The amount of information provided by the experiment \mathcal{E} , with prior knowledge $p(\theta)$, when the observation is x , is

$$(6) \quad s(\mathcal{E}, p(\theta), x) = s_1(x) - s_0.$$

This expression is also not invariant under a change of description of the parameter space.

The quantity $s(\mathcal{E}, p(\theta), x)$ depends on x ; some results are more informative than others. However, since θ is regarded as a random variable, this quantity may be averaged with respect to x according to the probability density given by (1). Hence, we have

DEFINITION 2. The average amount of information provided by the experiment \mathcal{E} , with prior knowledge $p(\theta)$, is

$$(7) \quad s(\mathcal{E}, p(\theta)) = E_x[s_1(x) - s_0].$$

Alternative forms for $s(\mathcal{E}, p(\theta))$ are

$$(8) \quad E_x E_\theta \log \{p(\theta | x)/p(\theta)\} \quad (\text{from (3) and (5)}),$$

$$(9) \quad E_x E_\theta \log \{p(x | \theta)/p(x)\} \quad (\text{from (2)}),$$

and, in full, if $p(x, \theta)$ is the joint density for x and θ ,

$$(10) \quad \iint p(x, \theta) \log \{p(x, \theta)/p(x)p(\theta)\} dx d\theta.$$

The expression (10) shows the symmetry between x and θ and also exhibits the fact that $s(\mathcal{E}, p(\theta))$ is invariant under a 1 - 1 transformation of the parameter space, Θ . The expression occurs in Shannon's theory ([10], Section 24) for the rate of transmission of information along a channel.⁴

Yet another expression for $s(\mathcal{E}, p(\theta))$ which is useful in calculation is obtained by introducing the information operator, I , along with the expectation operator, E . For a density function $p(y)$, we define

$$I_y p(y) = \int p(y) \log p(y) dy.$$

It is easy to verify that

$$(11) \quad s(\mathcal{E}, p(\theta)) = E_\theta I_x p(x | \theta) - I_x E_\theta p(x | \theta).$$

4. The results that we now proceed to establish involve only the use of Bayes' theorem and the two facts that the logarithm of a product is the sum of the two logarithms (in the combination of equations (12) and (13) for example) and that the function $x \log x$ is convex (in Theorem 1). We shall often denote the average information by $s(\mathcal{E})$ when the particular prior distribution does not have to be stressed.

THEOREM 1. $s(\mathcal{E}) \geq 0$, with equality if, and only if, $p(x | \theta)$ does not depend on θ , except possibly in a null set for θ .

This follows immediately from a well-known inequality (see, for example, Hardy, Littlewood, and Pólya [5], Theorem 205) on writing

$$s(\mathcal{E}) = \iint f(x, \theta) \log f(x, \theta) \cdot p(x)p(\theta) dx d\theta,$$

where

$$f(x, \theta) = p(x, \theta)/p(x)p(\theta).$$

The inequality says that

$$s(\mathcal{E}) \geq \iint f(x, \theta)p(x)p(\theta) dx d\theta \cdot \log \left\{ \frac{\iint f(x, \theta)p(x)p(\theta) dx d\theta}{\iint p(x)p(\theta) dx d\theta} \right\}$$

⁴ In the particular case of the "experiment" involved in radar work, the above ideas are already contained in a paper by P. M. Woodward [12], and are repeated in [13]. The author is indebted to M. S. Bartlett for these references.

with equality if, and only if, $f(x, \theta)$ equals a constant, except possibly on a null set. The logarithm is zero.

The theorem says that, provided the density of x varies with θ , any experiment is informative, on the average. Note that $\mathcal{I}(\mathcal{E}, p(\theta), x)$ is not necessarily nonnegative. Although the expectation is positive, the experimental result may reduce the amount of information. This can happen when a "surprising" value of x occurs; granted the correctness of the experimental technique, the "surprise" may result in our being less sure about θ than before the experiment.

Suppose that the observations x in an experiment \mathcal{E} consist of a pair of observations x_1, x_2 . That is, every $x \in \mathbf{X}$ is an ordered pair (x_1, x_2) with $x_i \in \mathbf{X}_i$ ($i = 1, 2$). Let \mathcal{G}_i be the σ -field over \mathbf{X}_i induced from \mathcal{G} by the transformation $x_i = x_i(x)$, and let P_i be the set of probability densities $p(x_i | \theta)$ of the observations x_i ($i = 1, 2$). (It is again supposed that the measures are, for all θ , dominated by a measure so that the probability distributions can be characterized by densities.) Then, $\mathcal{E}_i = \{\mathbf{X}_i, \mathcal{G}_i, \theta, P_i\}$ ($i = 1, 2$) are two experiments and \mathcal{E} is said to be the sum of the experiments \mathcal{E}_1 and \mathcal{E}_2 , written $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$. We shall also have to consider the experiment $\mathcal{E}_2(x_1) = \{\mathbf{X}_2, \mathcal{G}_2, \theta, P_2(x_1)\}$, where $P_2(x_1)$ is the set of densities $p(x_2 | \theta, x_1)$.

Consider $\mathcal{I}(\mathcal{E}_2(x_1), p(\theta | x_1))$. Since $p(\theta | x_1)$ is the posterior distribution of θ after x_1 has been observed, this quantity is the average information provided by an observation on x_2 after \mathcal{E}_1 has been performed and x_1 observed. The average of it over x_1 is defined to be the average information provided by \mathcal{E}_2 after \mathcal{E}_1 has been performed. We write it $\mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1)$, again suppressing $p(\theta)$. A proof along the lines of that for Theorem 1 establishes that $\mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) \geq 0$, with equality if, and only if, $p(x_2 | \theta, x_1)$ does not involve θ , except possibly on a null set.

THEOREM 2. $\mathcal{I}(\mathcal{E}_1) + \mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) = \mathcal{I}(\mathcal{E})$.

We have, using the form (9),

$$\begin{aligned} \mathcal{I}(\mathcal{E}_1) &= E_{x_1} E_{\theta} \log \{p(x_1 | \theta) / p(x_1)\} \\ (12) \quad &= E_{x_1} E_{x_2} E_{\theta} \log \{p(x_1 | \theta) / p(x_1)\}. \end{aligned}$$

Also, from the definitions immediately before the statement of the theorem,

$$\begin{aligned} \mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) &= E_{x_1} [\mathcal{I}(\mathcal{E}_2(x_1), p(\theta | x_1))] \\ (13) \quad &= E_{x_1} E_{x_2} E_{\theta} \log \{p(x_2 | \theta, x_1) / p(x_2 | x_1)\}. \end{aligned}$$

Addition of (12) and (13) gives

$$E_{x_1} E_{x_2} E_{\theta} \log \left\{ \frac{p(x_2 | \theta, x_1) p(x_1 | \theta)}{p(x_2 | x_1) p(x_1)} \right\} = E_{x_1} E_{x_2} E_{\theta} \log \left\{ \frac{p(x_1, x_2 | \theta)}{p(x_1, x_2)} \right\},$$

which is $\mathcal{I}(\mathcal{E})$, and the theorem is proved.

COROLLARY. If x_1 is sufficient for x in the Neyman-Fisher sense, then $\mathcal{I}(\mathcal{E}_1) = \mathcal{I}(\mathcal{E})$.

For if x_1 is sufficient for x , the factorization theorem shows that $p(x_2 | \theta, x_1)$

does not involve θ . Hence, by the remark immediately before the statement of the theorem, $\mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) = 0$, and the corollary is established.

The corollary establishes that there is no loss in information if attention is confined to observation on a sufficient statistic. Conversely, if a statistic is considered which is not sufficient (in the sense that it does not satisfy the factorization theorem), then information will be lost since $\mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) > 0$. Theorem 2 generalizes to a finite number of experiments with common Θ in an obvious manner.

DEFINITION 3. Two experiments, \mathcal{E}_1 and \mathcal{E}_2 , with $\Theta_1 = \Theta_2 = \Theta$, are independent if $p(x_1, x_2 | \theta) = p(x_1 | \theta)p(x_2 | \theta)$ for all $\theta \in \Theta$.

Of course it by no means follows that if \mathcal{E}_1 and \mathcal{E}_2 are independent, then x_1 and x_2 are independent; i.e., it is not usually true that $p(x_1, x_2) = p(x_1)p(x_2)$.

If \mathcal{E}_1 and \mathcal{E}_2 are independent, the experiments $\mathcal{E}_2(x_1)$ and \mathcal{E}_2 , defined above, are equivalent (in the sense that the four pairs of defining elements are all equal when we write $\mathcal{E}_1 \equiv \mathcal{E}_2$), and we have the result

$$(14) \quad \mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) = E_{x_1} \mathcal{I}(\mathcal{E}_2, p(\theta | x_1)).$$

THEOREM 3. If \mathcal{E}_1 and \mathcal{E}_2 are independent

$$\mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) \leq \mathcal{I}(\mathcal{E}_2),$$

with equality if, and only if, x_1 and x_2 are independent.

From (13) and the independence, we have

$$\begin{aligned} \mathcal{I}(\mathcal{E}_2) - \mathcal{I}(\mathcal{E}_2 | \mathcal{E}_1) &= E_{x_2} E_{\theta} \log \{p(x_2 | \theta)/p(x_2)\} \\ &\quad - E_{x_1} E_{x_2} E_{\theta} \log \{p(x_2 | \theta)/p(x_2 | x_1)\} \\ &= E_{x_1} E_{x_2} E_{\theta} \log \{p(x_2 | x_1)/p(x_2)\} \\ &= E_{x_1} E_{x_2} \log \{p(x_2 | x_1)/p(x_2)\}. \end{aligned}$$

The last expression is identical with (9) when x_2, x_1 are replaced by x, θ , respectively. By Theorem 1 it is therefore nonnegative, and is zero if, and only if, $p(x_2 | x_1) = p(x_2)$.

Again, the definition and theorem could be generalized to any finite number of independent experiments. The theorem says that if \mathcal{E}_1 and \mathcal{E}_2 are independent experiments, either one is more informative, on the average, if performed first than if performed second. In particular, if $\mathcal{E}_1 \equiv \mathcal{E}_2$, the theorem says that an independent repeat of the same experiment is less informative, on the average, than the original experiment. This is a property which agrees with the common belief in the diminishing marginal utility of independent equidistributed observations.

COROLLARY. If \mathcal{E}_1 and \mathcal{E}_2 are independent experiments, then

$$\mathcal{I}(\mathcal{E}_1) + \mathcal{I}(\mathcal{E}_2) \geq \mathcal{I}(\mathcal{E}),$$

with equality if, and only if, x_1 and x_2 are independent.

For

$$\begin{aligned} \mathcal{I}(\varepsilon_1) + \mathcal{I}(\varepsilon_2) &\geq \mathcal{I}(\varepsilon_1) + \mathcal{I}(\varepsilon_2 | \varepsilon_1) \quad (\text{by the theorem}) \\ &= \mathcal{I}(\varepsilon) \quad (\text{by Theorem 2}). \end{aligned}$$

The corollary is not necessarily true for experiments which are not independent. It is easy to construct an example where ε_1 or ε_2 separately provide no information, but jointly they are completely informative in the sense that the posterior distribution is necessarily concentrated on a single value of θ .

In the case of repetition of identical experiments, more than the result of Theorem 3 can be said about the reduction of information on repetition. Let $\varepsilon^{(1)} \equiv \varepsilon_1$ be any experiment and let $\varepsilon_2, \varepsilon_3, \dots$ be independent identical experiments. Let $\varepsilon^{(2)} = (\varepsilon_1, \varepsilon_2)$ and generally $\varepsilon^{(n)} = (\varepsilon_n, \varepsilon^{(n-1)})$. Let $\mathcal{I}(\varepsilon^{(n)}) = j_n$; the prior distribution can remain unspecified.

THEOREM 4. j_n is a concave, increasing function of n .

It will be enough to establish that

$$0 \leq j_{n+1} - j_n \leq j_n - j_{n-1}.$$

The first inequality follows from Theorem 2, for by that theorem

$$j_{n+1} - j_n = \mathcal{I}(\varepsilon_{n+1} | \varepsilon^{(n)}) \geq 0.$$

The second reads:

$$\mathcal{I}(\varepsilon_{n+1} | \varepsilon^{(n)}) \leq \mathcal{I}(\varepsilon_n | \varepsilon^{(n-1)}).$$

Since $\varepsilon_n \equiv \varepsilon_{n+1}$, it will be enough to show that

$$\mathcal{I}(\varepsilon_{n+1} | \varepsilon^{(n-1)}, \varepsilon_n) \leq \mathcal{I}(\varepsilon_{n+1} | \varepsilon^{(n-1)}).$$

This follows as a slight generalization of Theorem 3, saying that the additional experiment ε_n reduces the average information provided by ε_{n+1} , even after $\varepsilon^{(n-1)}$.

Consider the following experiment: With probability λ (for all values of θ), perform experiment ε_1 ; with probability $1 - \lambda$ (for all values of θ), perform ε_2 , where ε_1 and ε_2 have $\theta_1 = \theta_2 = \theta$. The observation will consist in the observation obtained, according to whichever experiment is performed, and the knowledge of which experiment was performed. Denote this experiment by $(\lambda\varepsilon_1 + (1 - \lambda)\varepsilon_2)$. In mathematical terms $(\lambda\varepsilon_1 + (1 - \lambda)\varepsilon_2) = \{X = X_1 \cup X_2, \mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2, \theta, P\}$, where P is the set of densities, $p(x | \theta)$, defined as follows: If $x \in X_1$, then $p(x | \theta) = \lambda p(x_1 | \theta)$ with $x = x_1$; if $x \in X_2$, then $p(x | \theta) = (1 - \lambda)p(x_2 | \theta)$ with $x = x_2$. It is easy to verify that

$$(15) \quad \mathcal{I}(\lambda\varepsilon_1 + (1 - \lambda)\varepsilon_2) = \lambda\mathcal{I}(\varepsilon_1) + (1 - \lambda)\mathcal{I}(\varepsilon_2).$$

In this terminology the concavity property established in Theorem 4 says that

$$\mathcal{I}(\lambda\varepsilon^{(k)} + (1 - \lambda)\varepsilon^{(m)}) \leq \mathcal{I}(\varepsilon^{(n)}),$$

with $n = \lambda k + (1 - \lambda)m$. The last equality ensures that the average "sample sizes" are the same, and the inequality says that rather than "mixing" two sample sizes, it is better to take a sample of fixed "size" equal to the average size of the mixture. We discuss the result again below.

THEOREM 5. For fixed \mathcal{E} , $g(\mathcal{E}, p(\theta))$ is a concave function of $p(\theta)$.

We have to show that if $p_1(\theta)$ and $p_2(\theta)$ are two prior probability densities and $0 \leq \lambda \leq 1$, then

$$g(\mathcal{E}, \lambda p_1(\theta) + (1 - \lambda)p_2(\theta)) - \lambda g(\mathcal{E}, p_1(\theta)) - (1 - \lambda)g(\mathcal{E}, p_2(\theta)) \geq 0.$$

The left-hand side is

$$\begin{aligned} & \iint p(x | \theta) (\lambda p_1(\theta) + (1 - \lambda)p_2(\theta)) \log \{p(x | \theta)/p(x)\} dx d\theta \\ & - \lambda \iint p(x | \theta) p_1(\theta) \log \{p(x | \theta)/p_1(x)\} dx d\theta \\ & - (1 - \lambda) \iint p(x | \theta) p_2(\theta) \log \{p(x | \theta)/p_2(x)\} dx d\theta, \end{aligned}$$

where $p_i(x) = \int p(x | \theta) p_i(\theta) d\theta$ ($i = 1, 2$) and $p(x) = \lambda p_1(x) + (1 - \lambda)p_2(x)$. This simplifies to give

$$\begin{aligned} & \lambda \iint p(x | \theta) p_1(\theta) \log \{p_1(x)/p(x)\} dx d\theta \\ & + (1 - \lambda) \iint p(x | \theta) p_2(\theta) \log \{p_2(x)/p(x)\} dx d\theta. \end{aligned}$$

Performing the integrations with respect to θ , we have

$$\lambda \int p_1(x) \log \{p_1(x)/p(x)\} dx + (1 - \lambda) \int p_2(x) \log \{p_2(x)/p(x)\} dx,$$

and these integrals are positive by the inequality used to establish Theorem 1.

THEOREM 6. Let $\mathcal{E}_i = \{\mathbf{X}, \mathcal{B}, \Theta, P_i\}$ ($i = 1, 2$). Let $\mathcal{E} = \{\mathbf{X}, \mathcal{B}, \Theta, P\}$, where P is the set of densities

$$p(x | \theta) = \lambda p_1(x | \theta) + (1 - \lambda)p_2(x | \theta),$$

with $0 \leq \lambda \leq 1$. Then

$$(16) \quad g(\mathcal{E}) \leq \lambda g(\mathcal{E}_1) + (1 - \lambda)g(\mathcal{E}_2).$$

(An alternative statement of this theorem reads: For fixed \mathbf{X} , \mathcal{B} , Θ , and $p(\theta)$, $g(\mathcal{E})$ is a convex function of P .)

The experiment \mathcal{E} , described in the statement of the theorem, can be thought of as being performed as follows: With probability λ , a value x is obtained according to the density $p_1(x | \theta)$; with probability $1 - \lambda$, x is obtained according to $p_2(x | \theta)$. The experimenter is informed only of x and not of which event,

of probability λ or $1 - \lambda$, took place. Let the experiment \mathcal{E}^* , on the other hand, inform him about this event but not about the value of x . Then, clearly, using the notation developed above,

$$(\mathcal{E}, \mathcal{E}^*) \equiv (\lambda \mathcal{E}_1 + (1 - \lambda) \mathcal{E}_2).$$

Hence,

$$g(\mathcal{E}) + g(\mathcal{E}^* | \mathcal{E}) = \lambda g(\mathcal{E}_1) + (1 - \lambda) g(\mathcal{E}_2)$$

and the result follows since $g(\mathcal{E}^* | \mathcal{E}) \geq 0$.

Note that we have a convexity property here and a concavity property in the previous theorem.

5. The previous development assigns to an experiment \mathcal{E} and a prior distribution $p(\theta)$ a numerical measure of the average information provided by \mathcal{E} . In particular, this permits a comparison to be made between the amounts of information provided by any two experiments $\mathcal{E}_1, \mathcal{E}_2$, with the same Θ , with respect to a prior distribution. It also allows \mathcal{E}_1 and \mathcal{E}_2 to be compared absolutely, that is, without reference to a prior distribution, in certain cases. To do this we introduce

DEFINITION 4. Let $\mathcal{E}_1, \mathcal{E}_2$ be two experiments with $\Theta_1 = \Theta_2 = \Theta$. \mathcal{E}_1 is more informative than \mathcal{E}_2 if

$$(17) \quad g(\mathcal{E}_1, p(\theta)) \geq g(\mathcal{E}_2, p(\theta))$$

for all $p(\theta)$,⁵ and strict inequality holds for some $p(\theta)$. We write $\mathcal{E}_1 > \mathcal{E}_2$ or $\mathcal{E}_2 < \mathcal{E}_1$. If equality holds in (17) for all $p(\theta)$, we say \mathcal{E}_1 and \mathcal{E}_2 are equally informative and write $\mathcal{E}_1 = \mathcal{E}_2$. We write $\mathcal{E}_1 \leq \mathcal{E}_2$, or $\mathcal{E}_2 \geq \mathcal{E}_1$, to mean either $\mathcal{E}_1 < \mathcal{E}_2$ or $\mathcal{E}_1 = \mathcal{E}_2$.

There exist pairs of experiments for which neither $\mathcal{E}_1 \geq \mathcal{E}_2$ nor $\mathcal{E}_1 \leq \mathcal{E}_2$. The merits of such experiments can only be judged by reference to a prior distribution. An example is given in the discussion of the binomial dichotomy after Theorem 9, below.

THEOREM 7. If $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ are three experiments with the same Θ and if \mathcal{E}_3 is independent of both \mathcal{E}_1 and \mathcal{E}_2 , then $\mathcal{E}_1 > \mathcal{E}_2$ implies $(\mathcal{E}_1, \mathcal{E}_3) > (\mathcal{E}_2, \mathcal{E}_3)$.

For any $p(\theta)$, by Theorem 2

$$\begin{aligned} g(\mathcal{E}_1, \mathcal{E}_3) &= g(\mathcal{E}_3) + g(\mathcal{E}_1 | \mathcal{E}_3) \\ &= g(\mathcal{E}_3) + E_{x_3} g(\mathcal{E}_1, p(\theta | x_3)), \end{aligned}$$

by (14), since \mathcal{E}_1 and \mathcal{E}_3 are independent. But $\mathcal{E}_1 > \mathcal{E}_2$ implies, in particular, that

$$g(\mathcal{E}_1, p(\theta | x_3)) \geq g(\mathcal{E}_2, p(\theta | x_3))$$

⁵ That is, for all prior distributions not merely for all prior distributions which are dominated by a fixed measure.

for any x_2 . Consequently,

$$\begin{aligned} g(\varepsilon_1, \varepsilon_3) &\geq g(\varepsilon_3) + E_{x_2} g(\varepsilon_2, p(\theta | x_2)) \\ &= g(\varepsilon_3) + g(\varepsilon_2 | \varepsilon_3), \end{aligned}$$

again by (14), since ε_2 and ε_3 are independent. A further application of Theorem 2 establishes the result.

THEOREM 8. *If ε_i ($i = 1, 2, 3, 4$) are four experiments with the same Θ and if $\varepsilon_1 > \varepsilon_2$, $\varepsilon_3 > \varepsilon_4$, ε_1 is independent of ε_3 , and ε_2 of ε_4 , then $(\varepsilon_1, \varepsilon_3) > (\varepsilon_2, \varepsilon_4)$.*

Let x_i be the random variable observed in ε_i . Then, for any value of θ , x_1 is independent of x_3 and x_2 of x_4 . Consider a new set of random variables (y_1, y_2, y_3, y_4) , where, for any value of θ , y_i has the same density as x_i , y_1 is independent of y_3 , y_2 of y_4 , and, in addition, y_2 is independent of y_3 . Let ε'_i denote the experiment in which y_i is observed. Clearly, for any $p(\theta)$, $g(\varepsilon_i) = g(\varepsilon'_i)$ and

$$g(\varepsilon_1, \varepsilon_3) = g(\varepsilon'_1, \varepsilon'_3) \geq g(\varepsilon'_2, \varepsilon'_3) \geq g(\varepsilon'_2, \varepsilon'_4) = g(\varepsilon_2, \varepsilon_4).$$

Both inequalities follow from Theorem 7, and the result is established.

Two other methods of comparing experiments have been introduced. The first, due to Bohnenblust, Shapley, and Sherman, says that ε_1 is more informative than ε_2 if every loss function attainable with ε_2 is also attainable with ε_1 . The second, due to Blackwell, says that ε_1 is sufficient for ε_2 if an experimenter performing ε_1 can, by a random device, obtain a result equivalent to performing ε_2 . (For a precise definition of these two relations, see Blackwell [2].) To avoid confusion, we shall speak of the relation introduced here as "more informative (S)" and the relation in terms of loss as "more informative (B)". For the latter, following Blackwell, we write $\varepsilon_1 \supset \varepsilon_2$, and for ε_1 is sufficient for ε_2 , we write $\varepsilon_1 > \varepsilon_2$. We remark that Theorems 7 and 8 above are the same as two theorems of Blackwell's (see [4], p. 332) with $>$ replacing \supset . We now discuss the connections between these three relations.

THEOREM 9. *If ε_1 and ε_2 are two experiments with the same Θ , and if ε_1 is sufficient for ε_2 , then ε_1 is not less informative (S) than ε_2 . In other words, $\varepsilon_1 > \varepsilon_2$ implies $\varepsilon_1 \geq \varepsilon_2$.*

Let x_i be the random variable observed in ε_i ($i = 1, 2$). $\varepsilon_1 > \varepsilon_2$ implies that there exists a stochastic transformation of x_1 , say x'_2 , such that $x'_2 \in \mathbf{X}_2$ and x'_2 and x_2 are identically distributed for each $\theta \in \Theta$. Let ε'_2 be the experiment in which x'_2 is observed. Clearly, $g(\varepsilon_2) = g(\varepsilon'_2)$. Consider the experiment $\varepsilon = (\varepsilon_1, \varepsilon'_2)$. Then, x_1 is sufficient for (x_1, x'_2) in the Neyman-Fisher sense, and, hence, by the Corollary to Theorem 2, $g(\varepsilon) = g(\varepsilon_1)$. But by Theorem 2, $g(\varepsilon) \geq g(\varepsilon'_2)$; consequently, $g(\varepsilon_1) \geq g(\varepsilon_2)$ for all $p(\theta)$, as required.

Conditions are known under which the relations \supset and $>$ are equivalent (see, for example, Blackwell [3]). Under these conditions, it will follow from Theorem 9 that not less informative (B) implies not less informative (S). That the converse of these results is not true can be illustrated by an example.

Consider the case of a binomial dichotomy. Here \mathbf{X} contains two elements $(0, 1)$, Θ contains two elements with $0 \leq \theta_1 \leq \theta_2 \leq 1$ and $p(x = 1 | \theta_i) = \theta_i = 1 - p(x = 0 | \theta_i)$. This experiment will be denoted by $\mathcal{E}(\theta_1, \theta_2)$. Denote the prior distribution over (θ_1, θ_2) by $(\lambda, 1 - \lambda)$ with $0 \leq \lambda \leq 1$. It follows immediately from (11) that

$$(18) \quad \mathcal{J}[\mathcal{E}(\theta_1, \theta_2), \lambda] = S(\lambda\theta_1 + (1 - \lambda)\theta_2) - \lambda S(\theta_1) - (1 - \lambda)S(\theta_2),$$

where

$$S(\theta) = -\theta \log \theta - (1 - \theta) \log (1 - \theta).$$

Consider a fixed experiment $\mathcal{E}(p_1, p_2)$ and compare it with $\mathcal{E}(\theta_1, \theta_2)$ as θ_1 and θ_2 vary. To do this it is necessary to consider the right-hand side of (18) as a function of λ for (p_1, p_2) and for (θ_1, θ_2) : $\mathcal{E}(p_1, p_2) \geq \mathcal{E}(\theta_1, \theta_2)$ if, and only if,

$$\mathcal{J}[\mathcal{E}(p_1, p_2), \lambda] \geq \mathcal{J}[\mathcal{E}(\theta_1, \theta_2), \lambda]$$

for all λ . It does not seem possible to describe the results analytically, and we therefore content ourselves with summarizing the results of some computations in the case $p_1 = \frac{1}{4}$, $p_2 = \frac{3}{4}$. The discussion is carried out with reference to Fig. 1, where P is the point $(\frac{1}{4}, \frac{3}{4})$. It is known (see [2]) that the points (θ_1, θ_2) in the areas with horizontal hatching correspond to experiments which can be compared with $\mathcal{E}(p_1, p_2)$ by either the relation \supset or $>$, which are, in this case, identical. For points in the triangular area, $\mathcal{E}(\theta_1, \theta_2) \subset \mathcal{E}(p_1, p_2)$; for points in the quadrilateral, $\mathcal{E}(\theta_1, \theta_2) \supset \mathcal{E}(p_1, p_2)$; the remaining experiments are not comparable with $\mathcal{E}(p_1, p_2)$. Theorem 9 implies that the relation \supset may be replaced by $>$, but computation shows that the points in the areas with vertical hatching correspond to additional experiments which can be compared with $\mathcal{E}(p_1, p_2)$ by the relation $>$. Those adjacent to the triangular area have $\mathcal{E}(\theta_1, \theta_2) < \mathcal{E}(p_1, p_2)$ and those adjacent to the quadrilateral have $\mathcal{E}(\theta_1, \theta_2) > \mathcal{E}(p_1, p_2)$. The points in the unhatched areas correspond to experiments which cannot be compared by the relation $>$. The points in the area of vertical hatching show that the converse of Theorem 9 is false.

The smallness of the unhatched region is a satisfactory feature of the comparison by the relation $>$, for ideally all experiments would be comparable.

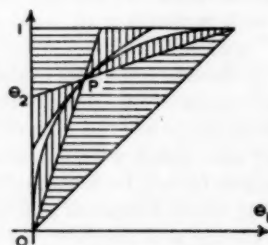


FIG. 1

The following considerations support the view that the relation $>$ holds in substantially more cases than does the relation \supset . Blackwell [2] remarks that the binomial trichotomy $\mathcal{E}_1 \equiv \mathcal{E}(0, \frac{1}{2}, 1)$, in an obvious extension of the previous notation, is not more informative than $\mathcal{E}_2 \equiv \mathcal{E}(0, \frac{1}{2}, \frac{1}{2})$. We shall show that $\mathcal{E}_1 > \mathcal{E}_2$.

Let $p(\theta) = \lambda p_1(\theta) + (1 - \lambda)p_2(\theta)$ and let $p_1(\theta) = (p, 0, q)$ and $p_2(\theta) = (p, q, 0)$. Then $p(\theta) = (p, q(1 - \lambda), q\lambda)$, a general prior distribution. From considerations of binomial dichotomies we have $\mathcal{J}(\mathcal{E}_1, p_2(\theta)) = \mathcal{J}(\mathcal{E}_2, p_2(\theta))$ and $\mathcal{J}(\mathcal{E}_1, p_1(\theta)) \geq \mathcal{J}(\mathcal{E}_2, p_1(\theta))$. From Theorem 5

$$\begin{aligned}\mathcal{J}(\mathcal{E}_1, p(\theta)) &\geq \lambda \mathcal{J}(\mathcal{E}_1, p_1(\theta)) + (1 - \lambda) \mathcal{J}(\mathcal{E}_1, p_2(\theta)) \\ &\geq \lambda \mathcal{J}(\mathcal{E}_2, p_1(\theta)) + (1 - \lambda) \mathcal{J}(\mathcal{E}_2, p_2(\theta)) \\ &= \lambda \mathcal{J}(\mathcal{E}_2, p(\theta)) + (1 - \lambda) \mathcal{J}(\mathcal{E}_2, p(\theta)) \\ &= \mathcal{J}(\mathcal{E}_2, p(\theta)).\end{aligned}$$

Since $p(\theta)$ is arbitrary and the inequality is clearly strict for some $p(\theta)$, the result is established.

Another difference between the two methods of comparison is provided by our Theorem 4. We deduced a result which we can now express as

$$(\lambda \mathcal{E}^{(k)} + (1 - \lambda) \mathcal{E}^{(m)}) \leq \mathcal{E}^{(n)},$$

where $n = \lambda k + (1 - \lambda)m$. W. H. Kruskal has pointed out that the same result is not necessarily true with \leq replaced by \subset . His example involves taking \mathcal{E}_1 to be a normal dichotomy, i.e., $\Theta = (\theta_1, \theta_2)$, \mathbf{X} is the real line, and

$$p(x | \theta_i) = (2\pi)^{-1/2} \exp[-\frac{1}{2}(x - \theta_i)^2].$$

Let $d_i (i = 1, 2)$ be two decisions, with d_i correct when $\theta = \theta_i$. Let $p_{ijn}(\delta)$ be the probability of saying d_i when $\theta = \theta_j$ on the evidence of the experiment $\mathcal{E}^{(n)}$, using the decision function δ . The relation \supset can be expressed in terms of $p_{12n}(\delta)$ and $p_{21n}(\delta)$, and for some values of c , the function

$$\inf_{\delta} \{p_{12n}(\delta) + cp_{21n}(\delta)\}$$

is not concave. Thus it may, to quote an extreme case, produce a smaller loss to do no experimentation with probability $(1 - \lambda)$ and to perform $\mathcal{E}^{(k)}$ with probability λ than to do $\mathcal{E}^{(n)}$ with $n = \lambda k$.

We conclude this section by discussing another example of Blackwell's (see [2]) which demonstrates the techniques of the present theory. Each member of a large population of individuals has or has not each of two characteristics H, S . The proportions, h and s , of individuals with characteristics H, S are known. The proportion w of individuals having both characteristics is not known. Let $\mathcal{E}(H)$ denote the experiment in which a random individual from the population of individuals having characteristic H is observed; use $\mathcal{E}(\sim H)$, $\mathcal{E}(S)$, and $\mathcal{E}(\sim S)$ similarly, where $\sim H$ denotes the absence of the characteristic H . Suppose, with-

out loss of generality, that the characteristics are so named that $0 \leq h \leq s \leq 1 - s \leq 1 - h \leq 1$. We proceed to show that $\mathcal{E}(H)$ is not less informative (S) than any of the other three experiments; that is, the best experiment is that in which individuals with the rarest characteristic are observed. Blackwell established the same result for not less informative (B) when w is known to be either hs or some specific alternative $\delta \neq hs$. Our result holds for any prior distribution of w .

Each of the four experiments is binomial with the following probabilities attached:

$$\mathcal{E}(H): \quad pr(S) = w/h = \theta,$$

$$\mathcal{E}(\sim H): \quad pr(\sim S) = \frac{1 - h - s + w}{1 - h} = \frac{1 - h - s}{1 - h} + \frac{h}{1 - h} \theta,$$

$$\mathcal{E}(S): \quad pr(H) = w/s = h\theta/s,$$

$$\mathcal{E}(\sim S): \quad pr(\sim H) = \frac{1 - h - s + w}{1 - s} = \frac{1 - h - s}{1 - s} + \frac{h}{1 - s} \theta,$$

where $\theta = w/h$. The permissible range for θ is $0 \leq \theta \leq 1$. Consider an arbitrary prior distribution for θ .

Now each of the four experiments is binomial with probability of the form $\lambda c + (1 - \lambda)\theta$, with $0 \leq \lambda \leq 1$, $0 \leq c \leq 1$. Alternatively, by introducing a random variable which is 1 or 0 according as the event, indicated above, does or does not occur, the probability density is

$$\begin{aligned} p(1 | \theta) &= \lambda c + (1 - \lambda)\theta \\ &= \lambda p_1(1 | \theta) + (1 - \lambda)p_2(1 | \theta), \end{aligned}$$

where $p_1(1 | \theta) = c$, $p_2(1 | \theta) = \theta$. Let $\mathcal{E}_1, \mathcal{E}_2$ be experiments with $P_1 = \{p_1\}$, $P_2 = \{p_2\}$. Then if \mathcal{E} is any of the four experiments considered above, we have by Theorem 6

$$g(\mathcal{E}) \leq \lambda g(\mathcal{E}_1) + (1 - \lambda)g(\mathcal{E}_2) = (1 - \lambda)g(\mathcal{E}_2) \leq g(\mathcal{E}_2),$$

since $g(\mathcal{E}_1) = 0$ as p_1 does not depend on θ . But $\mathcal{E}(H)$ has $\lambda = 0$, so that $\mathcal{E}_2 \equiv \mathcal{E}_H$. This establishes the result, since the prior distribution is arbitrary.

6. Since Wald's introduction of decision theory, many statisticians, the present author included, have identified the theory with statistical theory and have argued that modern statistics is decision theory. Some statisticians, for example, Barnard [1] and Fisher [15], have not supported this view; they have contended, for example, that the purpose of a significance test is different from the purpose of a Wald decision problem with two decisions, reject or accept. It is therefore contended that different mathematical models are needed for the two purposes. This latter view is supported by the fact that significance levels do not occur in decision theory. If the purpose of modern statistics is not to come

to decisions, we may ask what is its purpose? Without wishing to take sides in the issue we propose in this section of the paper to investigate some elementary consequences of the attitude that the purpose of *some* statistical experimentation is to gain and measure information about the state of nature.

The first consequence of this attitude is that the statistician, faced with a choice of one among several experiments that he might perform, will choose that one for which the average amount of information is the greatest. The choice will, in general, depend on his prior knowledge, but it may happen that the experiments will be absolutely comparable by the methods of Section 5 and the prior knowledge will be irrelevant. Examples have already been given, but there is one further case worth considering. Let $\mathcal{E}(\sigma)$ denote the experiment in which \mathbf{X} and Θ are the real lines and

$$p(x | \theta) = (\sqrt{2\pi}\sigma)^{-1} \exp [-(x - \theta)^2/2\sigma^2],$$

where $\sigma > 0$. Here x is normally distributed about θ with known variance σ^2 . We shall show that $\mathcal{E}(\sigma_1) > \mathcal{E}(\sigma_2)$ if $\sigma_1 < \sigma_2$; that is, the experiment with smaller variance is the more informative (S). To prove the result, we show that $\mathcal{E}(\sigma_1) > \mathcal{E}(\sigma_2)$ and then apply Theorem 9, with the additional remark that there obviously exists a $p(\theta)$ such that $\mathcal{J}(\mathcal{E}(\sigma_1), p(\theta)) > \mathcal{J}(\mathcal{E}(\sigma_2), p(\theta))$. Let x_i be the random variable observed in $\mathcal{E}(\sigma_i)$. Then

$$(19) \quad x'_2 = x_1 + u,$$

(where u is a random variable, independent of x_1 , and having a normal distribution with zero mean and variance $\sigma_2^2 - \sigma_1^2$) has, for each θ , the same distribution as x_2 . Equation (19) is thus a stochastic transformation from x_1 to x_2 and hence $\mathcal{E}(\sigma_1) > \mathcal{E}(\sigma_2)$.

A measure of the information can only be provided by assuming a particular form for $p(\theta)$. Suppose that

$$p(\theta) = (\sqrt{2\pi}\tau)^{-1} \exp [-(\theta - \mu)^2/2\tau^2] \equiv p,$$

for some μ and $\tau > 0$. It is easy to establish that $p(x)$ is a normal density with mean μ and variance $\sigma^2 + \tau^2$. Also,

$$I_\theta p(\theta) = -\log (2\pi e)^{1/2} \tau.$$

Consequently, by equation (11), we have

$$\mathcal{J}(\mathcal{E}(\sigma), p) = \frac{1}{2} \log (1 + \tau^2/\sigma^2).$$

This result provides an illustration of the truth of Theorem 4. If we use the notation of that theorem, with $\mathcal{E}_1 = \mathcal{E}(\sigma)$, we have that

$$j_n = \frac{1}{2} \log (1 + n\tau^2/\sigma^2),$$

which can be contrasted with the usual measure n/σ^2 . Notice that j_n increases without limit in this situation.

Consider now the k -dimensional extension of these results. Let \mathbf{X} and Θ be

k -dimensional Euclidean spaces and let $x = \{x_1, \dots, x_k\}$ have a multivariate normal density with mean $\theta = \{\theta_1, \dots, \theta_k\}$ and dispersion matrix C , which is known. Let θ have a prior density, p_A , which is multivariate normal with mean μ and dispersion matrix A . Denote this experiment by $\mathcal{E}(C)$; then, calculation along the same lines as in the univariate case gives

$$s(\mathcal{E}(C), p_A) = \frac{1}{2} \log \{|A + C| / |C|\}$$

where $|C|$ is the determinant of C . Clearly, even for this limited class of prior distributions, the two experiments $\mathcal{E}(C_1)$ and $\mathcal{E}(C_2)$ will not be absolutely comparable since their relative average informations depend critically on A . However, in some circumstances there is a possible simplification. Generally, we have that

$$s(\mathcal{E}(C_1), p_A) > s(\mathcal{E}(C_2), p_A)$$

if

$$|A + C_1| |C_2| > |A + C_2| |C_1|$$

or

$$|1 + A^{-1}C_1| |C_2| > |1 + A^{-1}C_2| |C_1|.$$

If the elements of $A^{-1}C_i$ are small in comparison with the unit matrix, this is approximately

$$|C_2| > |C_1|.$$

Hence, an approximate basis of comparison in this case, which corresponds to considerable ignorance about θ , is through the determinant of the dispersion matrix. The use of the determinant criterion has been used by Wald [11] in a slightly different context.

A second consequence of the view that one purpose of statistical experimentation is to gain information will be that the statistician will stop experimentation when he has enough information. Such a sequential method does not involve considerations of risks or cost of experimentation, but does involve a statement of prior knowledge. We consider next the sequential methods that this idea results in, for some special cases. In each case we shall consider a sequence $\mathcal{E}_1, \mathcal{E}_2, \dots$ of independent, identical experiments which are to be performed until enough information about θ has been obtained. It is therefore a question of how much repetition of a given experiment should be performed.

We first take the dichotomy, with $\Theta = (\theta_1, \theta_2)$. \mathbf{X} , \mathfrak{A} , and P are quite general. Let δ be some preassigned number. Then experimentation will proceed; after n repetitions we shall have observations (x_1, x_2, \dots, x_n) and the amount of information will be

$$(20) \quad \sum_i p_n(\theta_i) \log p_n(\theta_i),$$

where

$$p_n(\theta_i) = p(\theta_i | x_1, \dots, x_n),$$

the posterior distribution of θ . According to the idea introduced above, experimentation will continue until (20) is not less than δ . It is supposed that δ is chosen so that this sequential scheme will terminate with probability one; in this case δ must be negative. Since (20) is a convex function of $p_n(\theta_1) = 1 - p_n(\theta_2)$, the scheme corresponds to continuing sampling if, and only if,

$$(21) \quad 1 - A < p_n(\theta_1) < A,$$

where

$$A \log A + (1 - A) \log (1 - A) = \delta.$$

Expression (21) may be written in terms of the ratio of posterior probabilities for θ_1 and θ_2 , and by use of Bayes' theorem, it may be put in the form

$$\frac{1 - A}{A} \frac{p(\theta_2)}{p(\theta_1)} < \frac{p(x_1, \dots, x_n | \theta_1)}{p(x_1, \dots, x_n | \theta_2)} < \frac{A}{1 - A} \frac{p(\theta_2)}{p(\theta_1)}.$$

It is now apparent that the sampling scheme is equivalent to a scheme used in a Wald sequential probability ratio test of θ_1 against θ_2 .

The generalization to the case where Θ has n elements will be sufficiently illustrated by the trichotomy $n = 3$. The argument is as with the dichotomy up to the sentence before that in which (21) appears. Now the posterior distribution $p_n(\theta_i)$ may be represented by a point in an equilateral triangle of unit altitude, the distances of the point from the sides being $p_n(\theta_i)$ ($i = 1, 2, 3$). Since (20) is again a convex function of the distribution, it follows that for sufficiently large values of δ , but $\delta < 0$, the regions of values of $p_n(\theta_i)$ for which sampling will cease will be three congruent convex regions at the three corners of the triangle. The calculation of the exact shapes of the regions would be a simple matter. It is interesting to note that regions of similar convex structure are obtained for termination in an optimum sequential scheme for deciding between three simple hypotheses with given loss function and prior distribution (see, for example, Blackwell and Girshick [4], p. 262).

We now leave the case where Θ is finite and suppose Θ to be an interval on the real line. It is now necessary to remark that as Shannon's measure of information is not invariant under a change of description of the parameter space, a different sequential scheme will be obtained if the description is changed. This unpleasant feature need not bother us unduly since sequential schemes based, for example, on the variance will have a similar feature. A sampling scheme in which sampling is continued until the variance of the estimator of θ is less than some prescribed number will differ from one designed for the variance of the estimator of $f(\theta)$. It is possible to find invariant sequential schemes by the device of sampling until the *average* amount of information to be gained by taking a further sample falls below a prescribed limit. It can then be argued that the further sample is not worth taking and sampling can therefore cease. We shall not investigate such schemes here; they will be invariant since the expression for the average amount of information is invariant.

First consider repetitions of the normal experiment $\varepsilon(\sigma)$, above, with prior distribution p_τ . After n observations with mean \bar{x} , which is a sufficient statistic, it is easy to verify that the posterior distribution of θ is normal with mean $(n\tau^2\bar{x} + \sigma^2\mu) / (n\tau^2 + \sigma^2)$ and variance $\sigma^2\tau^2 / (n\tau^2 + \sigma^2)$. The posterior information will therefore be $-\frac{1}{2} \log 2\pi e \sigma^2 \tau^2 / (n\tau^2 + \sigma^2)$ and sampling will continue as long as this quantity is less than δ , or, equivalently, until

$$n \geq \frac{2\pi e \sigma^2 \tau^2 - \sigma^2 e^{-2\delta}}{\tau^2 e^{-2\delta}}.$$

Thus the optimum sequential scheme is of fixed sample size, given by the above expression. For large τ^2 , corresponding to small prior knowledge, the fixed sample size is approximately $n = K\sigma^2$, where $K = 2\pi e^{2\delta+1}$. Thus the scheme is equivalent to sampling until the variance of the sample mean is sufficiently small.

As a final example, consider the case of repeated binomial trials. In the experiment to be considered $\mathbf{X} = (0, 1)$, θ is the unit interval $0 \leq \theta \leq 1$, and $p(1 | \theta) = \theta$. The situation where the prior distribution is concentrated on a finite number of points is covered by the results above. We therefore consider densities over the whole interval of θ and, to simplify the calculations, confine attention to the family

$$(22) \quad p_{ab}(\theta) = \theta^{a-1}(1-\theta)^{b-1}\Gamma(a+b) / \Gamma(a)\Gamma(b),$$

with a and b positive. This family of densities has the property that if the prior distribution is $p_{ab}(\theta)$, then the posterior distribution after a single binomial trial has been performed is $p_{a+1,b}(\theta)$ or $p_{a,b+1}(\theta)$, according as $x = 1$ or 0 , respectively (a fact which the reader can easily verify). Simple calculation shows that

$$(23) \quad I_\theta p_{ab}(\theta) = \ln \Gamma(a+b) / \Gamma(a)\Gamma(b) + (a-1)[\Psi(a) - \Psi(a+b)] \\ + (b-1)[\Psi(b) - \Psi(a+b)],$$

where

$$\Psi(x) = d \ln \Gamma(x) / dx.$$

This complicated expression can be simplified for large values of both a and b by use of the asymptotic formulas

$$\ln \Gamma(x) \sim \frac{1}{2} \ln 2\pi - x + (x - \frac{1}{2}) \ln x$$

and

$$\Psi(x) \sim \ln x - 1/2x.$$

We obtain

$$I_\theta p_{ab}(\theta) \sim \frac{1}{2} \ln(a+b)^3 / ab - \frac{1}{2} \ln 2\pi - \frac{1}{2}.$$

It follows that the curve in the plane of a and b along which $I_\theta p_{ab}(\theta)$ is constant is given approximately, for large values of a and b , by the curve

$$(a+b)^3 = \lambda ab$$

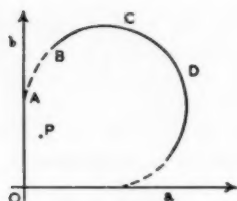


FIG. 2

for some constant λ . The general form of this curve is shown in Fig. 2 by a continuous line for $a, b > 10$; the broken extension shows the general form of the curve outside this range, as found by numerical computation.

Suppose the prior distribution has $a = a_0, b = b_0$. Then, after a sample of size n has produced r values of $x = 1$ and $n - r$ of $x = 0$, the posterior distribution will have $a = a_0 + r, b = b_0 + n - r$. The experimentation can be represented in the (a, b) -plane by starting at $P = (a_0, b_0)$ and forming a path by moving one unit along the a -axis for each value $x = 1$ and one unit along the b -axis for each value $x = 0$. Sampling will cease when the path intersects the curve corresponding to the amount of information required. If prior knowledge suggests that θ is small, then presumably one would take a_0 to be small and b_0 large, in comparison (for example, the point P in the figure). Ignorance about θ presumably corresponds to $a_0 = b_0 = 1$, or, at least a point with small a_0 and b_0 .

We conclude by making a few comments on the boundary curve shown in Fig. 2, based on the assumption that $a_0 = b_0 = 1$. The most prominent feature is perhaps the sharp decrease in the critical value of b as a approaches one—the curve AB in the figure. Repetitions of one value of x , in this case $x = 0$, result in a greater accumulation of information than a mixture of both values. To cite a numerical instance: 6 occurrences of the value $x = 0$ are about as informative as 11 occurrences of $x = 0$ with one occurrence of $x = 1$, or 14 of $x = 0$ with two of $x = 1$. (The sample sizes are 6, 12, and 16, respectively.) This agrees with the “common-sense” feeling adduced by the consideration that if the same thing continually happens, say the sun rises each morning, then we are much better informed than we would be if there was known to be even a single non-occurrence. In the contrary case, when θ is about $\frac{1}{2}$, the part CD of the curve is relevant, and is approximated to by the fixed sample size scheme with boundary $a + b = \text{constant}$. The part BC of the curve can also be approximated to by the straight-line boundary $b = \text{constant}$. This would be appropriate if θ were about $\frac{1}{2}$ (but not too small so that the sharp curve AB was relevant) and would correspond to sampling until b values of $x = 0$ had been observed. If $x = 1$ corresponds to a “defective,” this is the same as sampling, when defectives are rare, until the number of nondefectives has reached a preassigned number, and may be contrasted with inverse binomial sampling where the situation is similar but the rule is in terms of defectives.

7. Acknowledgments. I am much indebted to R. R. Bahadur for some valuable discussion, particularly in connection with Theorem 9, and to W. H. Kruskal for introducing me to the example in Section 5 and allowing me to include it in the present paper.

REFERENCES

- [1] G. A. BARNARD, "Simplified decision functions," *Biometrika*, Vol. 41 (1954), pp. 241-251.
- [2] DAVID BLACKWELL, "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 93-102.
- [3] DAVID BLACKWELL, "Equivalent comparisons of experiments," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 265-272.
- [4] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.
- [5] G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA, *Inequalities*, Cambridge University Press, 1952.
- [6] S. KULLBACK, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88-102.
- [7] S. KULLBACK, "Certain inequalities in information theory and the Cramér-Rao inequality," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 745-751.
- [8] S. KULLBACK AND R. A. LEIBLER, "Information and sufficiency," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 79-86.
- [9] BROCKWAY McMILLAN, "The basic theorems of information theory," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 196-219.
- [10] C. E. SHANNON, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379-423, 623-656.
- [11] ABRAHAM WALD, "On the efficient design of statistical investigations," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 134-140.
- [12] P. M. WOODWARD, "Theory of Radar Information" *Proceedings of a Symposium on Information Theory*, Ministry of Supply (London) (1950), pp. 108-113.
- [13] P. M. WOODWARD, *Probability and Information Theory, with Applications to Radar*, Pergamon Press Ltd., London, 1953.
- [14] LEE J. CRONBACH, "A consideration of information theory and utility theory as tools for psychometric problems" (Technical Report No. 1, Contract N6ori-07146, Urbana, Illinois, 1953).
- [15] R. A. FISHER, "Statistical methods and scientific induction," *J. Roy. Stat. Soc. (B)*, Vol. 17 (1955), pp. 69-78.

PROPERTIES OF SOME TWO-SAMPLE TESTS BASED ON A PARTICULAR MEASURE OF DISCREPANCY

BY L. H. WEGNER

The RAND Corporation

1. Introduction and summary. Let F and G be continuous univariate cdf's. For testing the hypothesis $F = G$ against general alternatives, E. Lehmann [4] has proposed and found certain properties of a test based on the particular measure of discrepancy $\int (F - G)^2 d[(F + G) / 2]$. In this note will be given some additional properties of Lehmann's test (cf. also [8]) and a closely related test proposed by Mood [2].

2. The test statistics. Let X_1, \dots, X_m and Y_1, \dots, Y_n be independent random samples from populations with continuous cdf's F and G respectively. Let $\binom{m}{2} \binom{n}{2} Q_{mn}$ be the number of quadruples (X_i, X_j, Y_k, Y_l) , $i < j, k < l$, for which either the maximum of the X 's is less than the minimum of Y 's or the maximum of the Y 's is less than the minimum of X 's. Then

$$(2.1) \quad Q_{mn} = \binom{m}{2}^{-1} \binom{n}{2}^{-1} \sum_{i=2}^m \sum_{k=1}^{i-1} \sum_{j=2}^n \sum_{l=1}^{j-1} X_{ijkl},$$

where X_{ijkl} is one if $X_i, X_j \leq Y_k, Y_l$ and is zero otherwise. Lehmann [4] has shown that Q_{mn} is a minimum variance unbiased estimate of the functional,

$$(2.2) \quad Q(F, G) = \frac{1}{3} + 2 \int (F - G)^2 d\left(\frac{F + G}{2}\right).$$

Replacing F and G in (2.2) by the corresponding sample cumulative distribution functions, say S and T , yields the statistic,

$$(2.3) \quad D = \int (S - T)^2 d\left(\frac{S + T}{2}\right),$$

which is the symmetric version of a test statistic originally proposed by Mood [2],

$$(2.4) \quad d = \int (S - T)^2 dT.$$

The critical region for each of the two-sample tests corresponding to the above test statistics consists of the region in the $m \times n$ dimensional sample space (or equivalently, the arrangements of the mX 's and nY 's) for which the test statistic takes on its largest values. In [8] the distribution of Q_{mn} when $F = G$ has been tabled for a selection of small sample sizes.

3. The two-sample statistics expressed in terms of ranks. In order to see

Received January 17, 1955.

more closely the similarities among the above statistics, it is enlightening to express them in terms of ranks. Let R_i be the rank of the i th ordered X and r_j the rank of the j th ordered Y in the combined ordered sample of mX 's and nY 's. Lehmann [4] has given the following relation between Q_{mn} and these ranks,

$$(3.1) \quad \binom{m}{2} \binom{n}{2} Q_{mn} = \sum_{j=1}^n \left[(n-j) \binom{r_j - j}{2} + (j-1) \binom{m - r_j + j}{2} \right].$$

From the definition of d , (2.4), we have

$$(3.2) \quad d = \frac{1}{n} \sum_{j=1}^n \left(\frac{j}{n} - \frac{r_j - j}{m} \right)^2,$$

and, by symmetry,

$$(3.3) \quad D = \frac{1}{2n} \sum_{j=1}^n \left(\frac{j}{n} - \frac{r_j - j}{m} \right)^2 + \frac{1}{2m} \sum_{i=1}^m \left(\frac{i}{m} - \frac{R_i - i}{n} \right)^2.$$

The above relations, after expansion and reduction, become

$$(3.4) \quad \begin{aligned} 2 \binom{m}{2} \binom{n}{2} Q_{mn} &= \sum_{j=1}^n \left[(n-1)r_j^2 - 2(m+n-2)jr_j \right. \\ &\quad \left. - (n-2m+1)r_j \right] + (n+2m-3) \frac{n(n+1)(2n+1)}{6} \\ &\quad + (n+m^2-3m+1) \frac{n(n+1)}{2} - mn(m-1), \end{aligned}$$

$$(3.5) \quad m^2 n^2 d = \sum_{j=1}^n [nr_j^2 - 2(m+n)jr_j] + (m+n)^2 \frac{(n+1)(2n+1)}{6},$$

$$(3.6) \quad \begin{aligned} 2m^2 n^2 D &= \sum_{j=1}^n [nr_j^2 - 2(m+n)jr_j] + \sum_{i=1}^m [mR_i^2 - 2(m+n)iR_i] \\ &\quad + (m+n)^2 \left[\frac{(n+1)(2n+1)}{6} + \frac{(m+1)(2m+1)}{6} \right]. \end{aligned}$$

4. Relations among tests when $m = n$. From the definition of Q_{mn} it follows that we may replace r_j by R_j in (3.4) if we interchange m and n . Upon adding the resultant expression for $2 \binom{m}{2} \binom{n}{2} Q_{mn}$ to (3.4), setting m equal to n , and employing the identities

$$\sum_{j=1}^n r_j + \sum_{j=1}^n R_j = n(2n+1)$$

and

$$\sum_{j=1}^n r_j^2 + \sum_{j=1}^n R_j^2 = \left(\frac{1}{3}\right)n(2n+1)(4n+1),$$

we obtain the following relation,

$$4 \binom{n}{2}^2 Q_{nn} = (n-1) \left[\frac{4}{3} n(2n+1)(n+1) + n^2(3n+1) - 4 \sum_{j=1}^n j(r_j + R_j) \right].$$

Proceeding in an analogous manner, we obtain from (3.6),

$$2n^4 D = n \left[\frac{1}{3} n(2n+1)(8n+5) - 4 \sum_{j=1}^n j(r_j + R_j) \right].$$

Thus, when $m = n$, Q_{nn} and D are related linearly and tests using large values of these statistics as critical regions are identical.

5. Means and variances.

a. *Means.* From (2.2) the mean of Q_{mn} is

$$(5.1) \quad E(Q_{mn}) = \frac{1}{3} + 2 \int (F - G)^2 d \left(\frac{F + G}{2} \right).$$

From (2.4),

$$(5.2) \quad E(d) = E \left[\int (S - T)^2 dT \right] = E \left(\int S^2 dT \right) - 2E \left(\int ST dT \right) + \frac{(n+1)(2n+1)}{6n^2}.$$

Since $E(T) = G$ and $E(S^2) = [F(1-F)/m] + F^2$, the first term on the right of (5.2) becomes, with the aid of Fubini's theorem,

$$E \left(\int S^2 dT \right) = \int \left[\frac{F(1-F)}{m} + F^2 \right] dG.$$

On the application of Fubini's theorem and a special integration by parts (*vide* [7], p. 102), the second term on the right of (5.2) reduces to

$$2E \left(\int ST dT \right) = \frac{n-1}{n} - \frac{n-1}{n} \int G^2 dF + \frac{2}{n} \int F dG.$$

Thus $E(d)$ may be written

$$E(d) = \frac{m-1}{m} \int F^2 dG + \frac{n-1}{n} \int G^2 dF + \frac{n-2m}{mn} \int F dG - \frac{n-1}{n} + \frac{(n+1)(2n+1)}{6n^2},$$

which, after substituting the identity,

$$\int F^2 dG + \int G^2 dF = \frac{1}{3} + \int (F - G)^2 d[(F + G)/2],$$

becomes

$$(5.3) \quad E(d) = \int (F - G)^2 d\left(\frac{F+G}{2}\right) - \frac{1}{m} \int F^2 dG - \frac{1}{n} \int G^2 dF \\ + \frac{n-2m}{mn} \int F dG + \frac{9n+1}{6n^2}.$$

From (5.3) and the symmetry of D , we obtain for $E(D)$,

$$(5.4) \quad E(D) = \int (F - G)^2 d\left(\frac{F+G}{2}\right) - \frac{1}{m} \int F^2 dG - \frac{1}{n} \int G^2 dF \\ + \frac{n-2m}{2mn} \int F dG + \frac{m-2n}{2mn} \int G dF + \frac{9n+1}{12n^2} + \frac{9m+1}{12m^2}.$$

When $F = G$, (5.1), (5.3), and (5.4) reduce to

$$(5.5) \quad E(Q_{mn}) = \frac{1}{3}$$

$$(5.6) \quad E(d) = \frac{1}{6} \left[\frac{m+n}{mn} + \frac{1}{n^2} \right]$$

$$(5.7) \quad E(D) = \frac{1}{6} \left[\frac{m+n}{mn} + \frac{1}{2n^2} + \frac{1}{2m^2} \right].$$

b. *Variances.* A method for finding the variance of Q_{mn} for general F and G has been given by Sundrum [8]. When $F = G$, he obtained

$$(5.8) \quad \sigma^2(Q_{mn}) = \frac{1}{45} \binom{m}{2}^{-1} \binom{n}{2}^{-1} \left[(m+n)(m+n-1) - 2 \right].$$

In the following there will be outlined a procedure (cf. Hoeffding [3]) for obtaining the variance of a U'_N statistic as defined in Theorem 6.5 (Q_{mn} is a particular U'_N statistic). This procedure also provides a result which will be needed in the proof of a later theorem.

Set

$$t_{ij}(x_1, \dots, x_i, y_1, \dots, y_j)$$

$$= Et(x_1, \dots, x_i, X_{i+1}, \dots, X_r, y_1, \dots, y_j, Y_{j+1}, \dots, Y_r),$$

$$\zeta_{00} = 0,$$

$$\zeta_{ij} = E[t_{ij}^2(X_1, \dots, X_i, Y_1, \dots, Y_j)] - \theta^2, \quad i, j = 0, \dots, r.$$

Let (s_1, \dots, s_r) , (s'_1, \dots, s'_r) , (t_1, \dots, t_r) , and (t'_1, \dots, t'_r) be four sets of r different integers, $1 \leq s_i, s'_i \leq m$, $1 \leq t_j, t'_j \leq n$, and let a and b be the number of integers common to the sets of s 's and t 's respectively. Then, from the symmetry of $t(x_1, \dots, x_r, y_1, \dots, y_r)$, it follows that

$$(5.9) \quad E[t(X_{s_1}, \dots, X_{s_r}, Y_{t_1}, \dots, Y_{t_r})t(X_{s'_1}, \dots, X_{s'_r}, Y_{t'_1}, \dots, Y_{t'_r})] - \theta^2 \\ = \zeta_{ab}.$$

Thus, the variance of U'_N can be written

$$\begin{aligned} \sigma^2(U'_N) &= \binom{m}{r}^{-2} \binom{n}{r}^{-2} E[\sum t(X_{s_1}, \dots, X_{s_r}, Y_{t_1}, \dots, Y_{t_r}) - \theta]^2 \\ (5.10) \quad &= \binom{m}{r}^{-2} \binom{n}{r}^{-2} \sum_{b=0}^r \sum_{a=0}^r \sum^{(ab)} E[t(X_{s_1}, \dots, X_{s_r}, Y_{t_1}, \dots, Y_{t_r}) \\ &\quad \cdot t(X_{s'_1}, \dots, X_{s'_r}, Y_{t'_1}, \dots, Y_{t'_r})] - \theta^2, \end{aligned}$$

where $\sum^{(ab)}$ stands for summation over all subscripts such that $1 \leq s_1 < \dots < s_r \leq m$, $1 \leq s'_1 < \dots < s'_r \leq m$, $1 \leq t_1 < \dots < t_r \leq n$, $1 \leq t'_1 < \dots < t'_r \leq n$ and exactly a equations $s_i = s'_i$ and b equations $t_k = t'_k$ are satisfied.

From (5.9) each term in $\sum^{(ab)}$ is equal to ζ_{ab} . The number of terms in $\sum^{(ab)}$ is $\binom{r}{a} \binom{m-r}{r-a} \binom{m}{r} \binom{n-r}{r-b} \binom{n}{r}$, so that (5.10) becomes

$$(5.11) \quad \sigma^2(U'_N) = \binom{m}{r}^{-1} \binom{n}{r}^{-1} \sum_{b=0}^r \sum_{a=0}^r \binom{m-r}{r-a} \binom{n-r}{r-b} \binom{r}{a} \binom{r}{b} \zeta_{ab}.$$

To find the variance of Q_{mn} by the above method, we must first obtain the ζ_{ab} 's $a, b = 0, 1, 2$, and then combine them according to (5.11).

Since $\zeta_{00} = 0$, we have the following additional result, which will be needed in the following section.

$$\begin{aligned} (5.12) \quad \sigma^2(U'_N) &\leq \binom{m}{r}^{-1} \binom{n}{r}^{-1} \left[\binom{m}{r} \binom{n}{r} - \binom{m-r}{r} \binom{n-r}{r} \right] \max(\zeta_{ab}), \\ &= o(1) \text{ as } \min(m, n) \rightarrow \infty, \end{aligned}$$

when $\max(\zeta_{ab}) < \infty$.

From (3.5) the variance of d is

$$\begin{aligned} (5.13) \quad \sigma^2(d) &= m^{-4} n^{-4} \left[n^2 \text{var} \left(\sum_{j=1}^n r_j^2 \right) + 4(m+n)^2 \text{var} \left(\sum_{j=1}^n j r_j \right) \right. \\ &\quad \left. - 4n(m+n) \text{cov} \left(\sum_{i=1}^n r_i^2, \sum_{j=1}^n j r_j \right) \right]. \end{aligned}$$

When $F = G$, the distribution of r_j and the joint distribution of r_j and r_k ($j \leq k$) are easily seen to be

$$(5.14) \quad f(r_j) = \binom{m+n}{n}^{-1} \binom{r_j-1}{j-1} \binom{m+n-r_j}{n-j}, \quad \begin{matrix} j \leq r_j \leq m+j, \\ 1 \leq j \leq n \end{matrix}$$

and

$$\begin{aligned} (5.15) \quad f(r_j, r_k) &= \binom{m+n}{n}^{-1} \binom{r_j-1}{j-1} \binom{r_k-r_j-1}{k-j-1} \\ &\quad \cdot \binom{m+n-r_k}{n-k}, \quad \begin{matrix} j \leq r_j \leq r_k - (k-j) \leq m+j, \\ 1 \leq j \leq k \leq n. \end{matrix} \end{aligned}$$

Equations (5.14) and (5.15) may be used to find the three terms on the right of (5.13). After lengthy but straightforward calculations, we have

$$(5.16) \quad \text{var} \left(\sum_{j=1}^n r_j^2 \right) = \frac{1}{180} mn(m+n+1)(2m+2n+1)(8m+8n+11),$$

$$(5.17) \quad \text{var} \left(\sum_{j=1}^n jr_j \right) = \frac{1}{180} mn(m+n+1)(2n+3)(2n+1),$$

$$(5.18) \quad \begin{aligned} & \text{cov} \left(\sum_{i=1}^n r_i^2, \sum_{j=1}^n jr_j \right) \\ &= \frac{1}{360} mn(m+n+1)(16n^2 + 16mn + 14m + 31n + 13). \end{aligned}$$

The following relation, which will be used in Theorem 6.1, can be obtained in a similar manner,

$$(5.19) \quad \text{var} \left(\sum_{j=1}^n r_j \right) = \frac{1}{12} mn(m+n+1).$$

Substituting (5.16), (5.17), and (5.18) in (5.13) and simplifying, we obtain for the variance of d when $F = G$,

$$(5.20) \quad \sigma^2(d) = \frac{(m+n)(m+n+1)}{45m^2n^2} + \frac{m+n+1}{180m^2n^3} (12m^2 - 3n^2 - 2mn).$$

6. Limiting distributions.

a. *Under the null hypothesis.* The following two theorems are concerned with the limiting distribution of Q_{mn} , d , and D under the null hypothesis $F = G$.

THEOREM 6.1. *If $F = G$ and $m/n \rightarrow c > 0$ as $n \rightarrow \infty$, the statistics*

$$\frac{1}{2} \frac{mn}{m+n} [Q_{mn} - E(Q_{mn})], \quad \frac{mn}{m+n} [D - E(D)], \quad \frac{mn}{m+n} [d - E(d)]$$

have the same limiting distribution.

THEOREM 6.2.¹ *If $F = G$ and $m/n \rightarrow c > 0$ as $n \rightarrow \infty$, the statistic $mn/(m+n)[d - E(d)]$ has the same limiting distribution as $n\omega^2 - E(n\omega^2)$, where ω^2 is the von Mises statistic. (The limiting distribution of $n\omega^2$ is tabled in [1]; $E(n\omega^2) = \frac{1}{6}$.)*

It follows from the above theorems that $\frac{1}{2}[(mn/m+n)][Q_{mn} - E(Q_{mn})]$ has the same limiting distribution as $n\omega^2 - E(n\omega^2)$. In Figure 1 are compared the distribution of Q_{55} and the limiting distribution of $n\omega^2$ drawn with the appropriate scales.

PROOF OF THEOREM 6.1. From equations (3.4) and (3.5), we may write

$$(6.1) \quad \begin{aligned} & 2 \binom{m}{2} \binom{n}{2} Q_{mn} - m^2 n^2 d^2 \\ &= - \sum_{j=1}^n r_j^2 + 4 \sum_{j=1}^n jr_j - (n-2m+1) \sum_{j=1}^n r_j + O(n^2). \end{aligned}$$

From (5.16), (5.17), and (5.19), each of the terms on the right has variance $O(n^2)$.

¹ Theorem 6.2 is due to M. Rosenblatt [6].

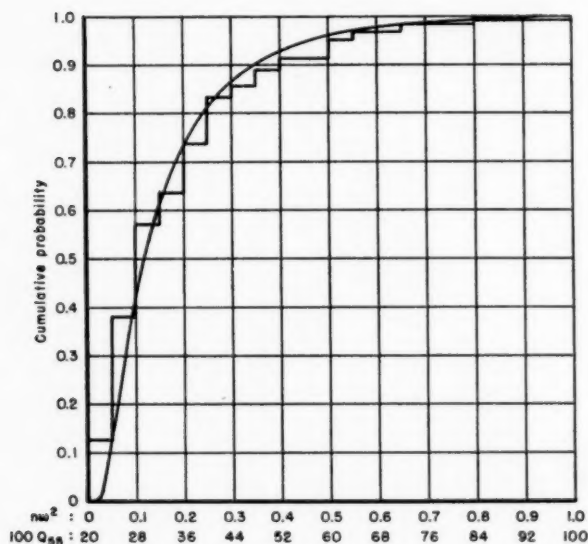


FIG. 1

Thus, aside from terms which converge to zero in probability,

$$(6.2) \quad \frac{1}{2} \frac{mn}{m+n} [Q_{mn} - E(Q_{mn})] = \frac{mn}{m+n} \frac{1}{4} \binom{m}{2}^{-1} \binom{n}{2}^{-1} m^2 n^2 [d - E(d)]$$

$$= \frac{mn}{m+n} [d - E(d)] + \frac{mn}{m+n} \frac{m+n-1}{(m-1)(n-1)} [d - E(d)].$$

From (5.20), $\sigma^2(d) = O(1/n^2)$. Thus the second term on the right of (6.2) converges to zero in probability.

The proof that $mn / (m+n) [D - E(D)]$ has the same limiting distribution as $\frac{1}{2} [mn / (m+n)] [Q_{mn} - E(Q_{mn})]$ is analogous and will be omitted.

b. *Under the alternative hypothesis.* An important subclass of the class of continuous cdf's is the class of strictly increasing continuous cdf's. The following two theorems are concerned with the problem of finding the limiting distribution of Q_{mn} , d , and D when F and G are in this subclass and $F \neq G$.

THEOREM 6.3. *If $m/n \rightarrow c > 0$ as $n \rightarrow \infty$, then the statistics*

$$\frac{1}{2} [mn / (m+n)]^{1/2} [Q_{mn} - E(Q_{mn})], \quad [mn / (m+n)]^{1/2} [d - E(d)], \quad \text{and} \\ [mn / (m+n)]^{1/2} [D - E(D)]$$

have the same limiting distribution.

PROOF. It follows from (6.1) and the inequalities,

$$\sum_{j=1}^n j r_j \leq \sum_{j=1}^n r_j^2 < n(m+n)^2 \quad \text{and} \quad \sum_{j=1}^n r_j < n(m+n),$$

that we may write, aside from terms which converge to zero in probability,

$$\begin{aligned}
 \sqrt{\frac{mn}{m+n}} [d - E(d)] &= \sqrt{\frac{mn}{m+n}} m^{-2} n^{-2} 2 \binom{m}{2} \binom{n}{2} [Q_{mn} - E(Q_{mn})] \\
 (6.3) \qquad &= \frac{1}{2} \sqrt{\frac{mn}{m+n}} [Q_{mn} - E(Q_{mn})] \\
 &\quad + \sqrt{\frac{mn}{m+n}} \frac{1-m-n}{2mn} [Q_{mn} - E(Q_{mn})].
 \end{aligned}$$

From (5.12) it follows that $\sigma^2(Q_{mn}) = o(1)$ as $\min(m, n) \rightarrow \infty$, so that the second term on the right of (6.5) converges to zero in probability.

The proof that $[mn / (m+n)]^{1/2} [D - E(D)]$ has the same limiting distribution as $\frac{1}{2} [mn / (m+n)]^{1/2} [Q_{mn} - E(Q_{mn})]$ is analogous and will be omitted.

THEOREM 6.4.² *If $m/n = c > 0$ as $n \rightarrow \infty$, then the statistic $[mn / (m+n)]^{1/2} [Q_{mn} - E(Q_{mn})]$ has a normal limiting distribution. Excluding F and G for which either $F = G$ or $\int F dG = 0$ or 1, the class for which nondegeneracy occurs includes all continuous F and G which are strictly increasing throughout their range of variation.*

In the proof of Theorem (6.4) we shall need the following theorem of Lehmann's [4].

THEOREM 6.5. *Let X_1, \dots, X_r , and Y_1, \dots, Y_n be independently distributed random samples from the distributions F and G respectively. Let $t(x_1, \dots, x_r, y_1, \dots, y_r)$ be symmetric in the x 's alone and in the y 's alone. Suppose that*

$$E[t(X_1, \dots, X_r, Y_1, \dots, Y_r)] = \theta(F, G) = \theta,$$

$$E[t(X_1, \dots, X_r, Y_1, \dots, Y_r)^2] = M < \infty.$$

Let $m/n = c$ and let n be sufficiently large so that $r \leq \min(m, n)$. Define

$$U'_n = \binom{m}{r}^{-1} \binom{n}{r}^{-1} \sum t(X_{\alpha_1}, \dots, X_{\alpha_r}, Y_{\beta_1}, \dots, Y_{\beta_r}),$$

where the summation is extended over all subscripts

$$1 \leq \alpha_1 < \dots < \alpha_r \leq m, \quad 1 \leq \beta_1 < \dots < \beta_r \leq n.$$

Then, as $n \rightarrow \infty$, $[mn / (m+n)]^{1/2} (U'_n - \theta)$ is asymptotically normally distributed; furthermore, if we set

$$\psi_1(x_1) = E[t(x_1 X_2, \dots, X_r, Y_1, \dots, Y_r)] - \theta,$$

$$\psi_2(y_1) = E[t(X_1, \dots, X_r, y_1 Y_2, \dots, Y_r)] - \theta,$$

then the limiting distribution of $[mn / (m+n)]^{1/2} (U'_n - \theta)$ is nondegenerate provided

$$E[\psi_1^2(X_1)] + E[\psi_2^2(Y_1)] > 0.$$

² Theorem 6.4 is an amended version of a statement by E. Lehmann [4, p. 173], which did not sufficiently restrict F and G for nondegeneracy.

PROOF OF THEOREM (6.4). Set $t(X_i, X_j, Y_k, Y_l)$ equal to X_{ijkl} [which is defined following (2.1)]. Then Q_{mn} is seen to be equivalent to U'_n in Theorem 6.5, where $\theta = Q(F, G)$ and $r = 2$. The first statement of Theorem 6.4 follows immediately. To prove the second statement, we apply the second part of Lehmann's theorem. We have

$$\begin{aligned}\psi_1(x_1) &= E[t(x_1, X_2, Y_1, Y_2)] - \theta \\ &= \int_{-\infty}^{x_1} \int_y^{\infty} 2G(y) dF(x) dG(y) + \int_{x_1}^{\infty} \int_{-\infty}^y 2[1 - G(y)] dF(x) dG(y) - \theta \\ &= 2 \int_{-\infty}^{x_1} (1 - F)G dG + 2 \int_{x_1}^{\infty} F(1 - G) dG - \theta \\ &= 2 \int_{-\infty}^{x_1} (G - F) dG + 2 \int_{x_1}^{\infty} F(1 - G) dG - \theta.\end{aligned}$$

Set $I(x_1) = \int_{-\infty}^{x_1} (G - F) dG$. Then $E[\psi_1^2(X_1)] = 0$ implies that

$$(6.4) \quad I(X_1) = E[I(X_1)]$$

with probability one (with respect to F).

Suppose now that the restrictions of the second statement of Theorem 6.4 hold. This implies that there exist two points x_0 and x'_0 , $x_0 < x'_0$, which are points of increase of both F and G . With no loss in generality it may be assumed that $G(x) - F(x) \geq \delta > 0$ for x in the interval (x_0, x'_0) . It follows that $I(x'_0) > I(x_0)$ so that (6.4) can not hold with probability one.

7. Consistency and unbiasedness.

a. *Consistency.* For the class of continuous cumulative distribution functions, the test based on Q_{mn} of the hypothesis $F = G$ against the alternatives $F \neq G$ has been shown by Lehmann [4] to be consistent at each level of significance when $\min(m, n) \rightarrow \infty$.

With the aid of the theorems on limiting distributions and the fact that the means of d and D are linear functions of $\int (F - G)^2 d(F + G)$ plus a term which is $o(1)$ as $\min(m, n) \rightarrow \infty$, it readily follows that the tests based on d and D are consistent under the conditions of the above paragraph provided the additional restriction $m/n \rightarrow c > 0$ as $n \rightarrow \infty$ is imposed.

b. *Unbiasedness.* That the tests based on Q_{mn} , d , and D are not unbiased tests of the hypothesis $F = G$ against all continuous alternatives $F \neq G$ and all m and n is shown by the following example.

Let F_1 and G_1 be cdf's with the probability density functions

$$\begin{aligned}f_1(x) &= \frac{1}{2}, & 0 \leq x \leq 1, 2 \leq x \leq 3 \\ &= 0, & \text{otherwise,}\end{aligned}$$

and

$$\begin{aligned}g_1(y) &= 1, & 1 \leq y \leq 2 \\ &= 0, & \text{otherwise.}\end{aligned}$$

In the $m \times n$ dimensional sample space let $W_{mn}^{(1)}$ be the region for which $\max(x_1, \dots, x_m) < \min(y_1, \dots, y_n)$ and let $W_{mn}^{(2)}$ be the region for which $\min(x_1, \dots, x_m) > \max(y_1, \dots, y_n)$. Then

$$P(W_{mn}^{(1)} | F = G) = P(W_{mn}^{(2)} | F = G) = \binom{m+n}{n}^{-1},$$

$$P(W_{mn}^{(1)} | F = F_1, G = G_1) = P(W_{mn}^{(2)} | F = F_1, G = G_1) = \left(\frac{1}{2}\right)^m.$$

Since, for fixed n and sufficiently large m , $\binom{m+n}{n} < 2^m$, there exist m_1 and n_1 such that both

$$P(W_{m_1 n_1}^{(1)} | F = F_1, G = G_1) < P(W_{m_1 n_1}^{(1)} | F = G),$$

$$P(W_{m_1 n_1}^{(2)} | F = F_1, G = G_1) < P(W_{m_1 n_1}^{(2)} | F = G),$$

so that any test of the hypothesis $F = G$ having $W_{m_1 n_1}^{(1)}$, $W_{m_1 n_1}^{(2)}$, or $W_{m_1 n_1}^{(1)} \cup W_{m_1 n_1}^{(2)}$ as a critical region will be biased against the alternative $F = F_1, G = G_1$.

Since critical regions for the tests based on Q_{mn} , d , and D are regions yielding large values of these statistics, it can be seen by examining the maxima of these statistics over the possible arrangements of X 's and Y 's that for each test and every m and n , one of $W_{mn}^{(1)}$, $W_{mn}^{(2)}$, or $W_{mn}^{(1)} \cup W_{mn}^{(2)}$ is a possible critical region. Thus, each of these tests is biased against the alternative $F = F_1, G = G_1$, when $m = m_1, n = n_1$.

8. The power of the test based on Q_{mn} for a particular class of alternatives.

In [5], Lehmann has discussed the power of several two-sample distribution-free tests for the particular class of alternatives $G = F^k (k = 2, 3, \dots)$. One of the tests considered by Lehmann was the two-sided version of Wilcoxon's rank sum test, which we shall use here as a basis of comparison. With the aid of Lehmann's technique, the exact power of the Q_{mn} test was found for $m = n = 4$ to be 0.19 against the alternative $G = F^2$ and 0.32 against $G = F^3$, which results are identical with the corresponding results for Wilcoxon's test. For larger m and n , the approximate power of the Q_{mn} test was obtained by use of the approximate distributions indicated in Section 6. Against the alternative $G = F^2$, the approximate power was slightly larger than that of Wilcoxon's test for $5 \leq m = n \leq 40$ and slightly smaller for $m = n > 40$. Against the alternative $G = F^3$, the approximate power was essentially the same as that of Wilcoxon's test for $5 \leq m = n < 15$ and slightly smaller for $m = n \geq 15$.

REFERENCES

- [1] T. W. ANDERSON AND D. A. DARLING, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193-213.
- [2] W. J. DIXON, "A criterion for testing the hypothesis that two samples are from the same population," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 199-204.
- [3] W. HOFFMANN, "A class of statistics with asymptotic normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293-325.

- [4] E. L. LEHMANN, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165-179.
- [5] E. L. LEHMANN, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 23-44.
- [6] M. ROSENBLATT, "Limit theorems associated with variants of the von Mises statistics," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 617-624.
- [7] S. SAKS, *Theory of the Integral*, New York, Hafner Publishing Co., 1937.
- [8] R. M. SUNDRUM, "On Lehmann's two-sample test," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 139-146.

POWER OF ANALYSIS OF VARIANCE TEST PROCEDURES FOR CERTAIN INCOMPLETELY SPECIFIED MODELS, I^{1, 2}

BY HELEN BOZIVICH,³ T. A. BANCROFT, AND H. O. HARTLEY

Statistical Laboratory, Iowa State College

TABLE OF CONTENTS

	PAGE
1. Introduction.....	1017
1.1 Description of pooling procedures.....	1017
1.2 More precise formulation.....	1019
1.3 Reduction of mixed models to random models.....	1021
1.4 Related papers and objectives of present investigation.....	1021
2. Exact and approximate formulas for power. Component of variance model.....	1022
2.1 Mathematical formulation of the pooling procedure.....	1022
2.2 Integral expressions for the power.....	1022
2.3 Exact formulas.....	1024
2.31 Series formulas.....	1024
2.32 Recurrence formulas.....	1025
2.4 Approximate formulas.....	1026
2.5 Theory of reduction of mixed model to random model.....	1027
2.6 Application of derived formulas.....	1028
3. Discussion of power and size and comparison of test procedures.....	1029
3.1 Type of recommendations attempted.....	1029
3.2 Size.....	1029
3.3 Frequency of pooling.....	1036
3.4 Power.....	1037
References cited.....	1042
Appendix: Figures and tables.....	1041

1. Introduction.

1.1 *Description of pooling procedures.* The simplest situation of a pooling procedure for testing hypotheses using analysis of variance procedures may be described as follows: We are given three mean squares, V_1 , V_2 , V_3 , based on n_1 , n_2 , and n_3 degrees of freedom, respectively, and designated as treatment mean square (V_3), the error mean square (V_2), and the doubtful error mean square (V_1). It is desired to test a null hypothesis involving V_3 , which can be tested by comparing V_3 with V_2 by the F -test. It is now suspected that V_1 is also a measure of the error variance, that is, has the same expectation as V_2 . It is decided, therefore, to first perform a preliminary test of significance by comparing

Received July 12, 1955; revised May 21, 1956.

¹ Journal paper No. J-2962 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 169.

² This paper is based on research sponsored by the Iowa State College Agricultural Experiment Station and the Wright Air Development Center.

³ Part of the results presented here are contained in a thesis prepared by Miss Helen Bozivich and submitted to the Graduate Faculty of Iowa State College in partial fulfillment of the requirements for the Ph.D. in Statistics, June, 1955.

V_2 against V_1 by the F -test, and, if this turns out to be nonsignificant, to use the pooled mean square $V = (n_1V_1 + n_2V_2)/(n_1 + n_2)$ as error for comparison with V_3 in the final F -test. In the case that V_2 is significantly different from V_1 , however, use V_2 as error in the final F -test. This test procedure is usually referred to as the sometimes-pool procedure. In corresponding terminology the single test V_3/V_2 is called the never-pool test, and the procedure of employing $V = (n_1V_1 + n_2V_2)/(n_1 + n_2)$ as error and always testing only V_3/V is called the always-pool test. If the level of significance for the preliminary test is 100%, the sometimes-pool test becomes the never-pool test; if, on the other hand, the level is 0%, the sometimes-pool test becomes the always-pool test. With the sometimes-pool test, the precise nature of the final F -test is, therefore, not determined in advance, but it depends on the relative magnitude of the observed mean squares V_2 and V_1 .

When the analysis of variance and associated tests of significance were first developed by R. A. Fisher, such procedures were not advocated. Indeed, Fisher's original description of analysis of variance tests clearly stipulated that for every well-designed experiment there can only be one correct analysis and the test(s) of significance are completely determined before the experimental results are available. With Fisher the appropriate test of significance is determined by a specification of the population from which the experimental data were sampled. We may speak in this case of an analysis determined by a completely specified model.⁴ However, in experimental design, situations frequently arise in which the model is not completely specified. Furthermore, with the wider application of analysis of variance to operational research and to the study of routine data, statisticians are often faced with analyzing data which have not resulted from a designed experiment, and in these situations the model is often incompletely specified. In such cases preliminary tests of significance have been used, in practice, as an aid in choosing an appropriate specification from which valid subsequent inferences may be drawn. In particular, preliminary tests of significance procedures have been used in the past in an attempt to increase the number of degrees of freedom associated with the error in a final F -test, thereby apparently increasing the sensitivity of the final F -test. Justification for the use of such methods has been made apparently on intuitive grounds.

The procedures described above and similar pooling procedures can be regarded as dealing with these situations sequentially in two stages: the preliminary stage in which inferences are drawn about the model, and the final stage in which inferences are drawn about the parameter(s) involved in the main hypothesis.

The purpose of the present study is to critically examine the consequences of certain pooling procedures with regard to the resulting errors of the first and second kind, for certain random and mixed models. Finally, on the basis of these results, we shall attempt some general recommendations on the advisability or otherwise of using them.

⁴ For the general formulation of model specification, see Section 1.2.

TABLE 1
Component of variance model with $\sigma_b^2 > 0$ —Analysis of variance

Source of variation	df	Mean square	Exp. mean square
Between <i>A</i>	$n_2 = q - 1$	V_2	$\sigma_a^2 + s\sigma_b^2 + r\sigma_s^2 = \sigma_a^2$
Between <i>B</i> within <i>A</i>	$n_2 = q(r - 1)$	V_2	$\sigma_a^2 + s\sigma_b^2 = \sigma_b^2$
Within <i>B</i>	$n_1 = qr(s - 1)$	V_1	$\sigma_s^2 = \sigma_s^2$

1.2 *More precise formulation.* Let us assume the component of variance model

$$(1) \quad x_{ijk} = \mu + a_i + b_{ij} + z_{ijk},$$

where $i = 1, 2, \dots, q$; $j = 1, 2, \dots, r$; $k = 1, 2, \dots, s$; a_i is $N(0, \sigma_a^2)$, b_{ij} is $N(0, \sigma_b^2)$, and z_{ijk} is $N(0, \sigma_s^2)$. We wish to test a hypothesis concerning a_i . If $\sigma_b^2 \geq 0$ and $\sigma_s^2 > 0$, then

$$(2) \quad \begin{aligned} x_{ijk} &= \mu + a_i + b_{ij} + z_{ijk} && \text{for } \sigma_b^2 > 0, \\ x_{ijk} &= \mu + a_i + z_{ijk} && \text{for } \sigma_b^2 = 0. \end{aligned}$$

In this case (1) is said to be an incompletely specified model. If, however, $\sigma_b^2 > 0$,

$$(3) \quad x_{ijk} = \mu + a_i + b_{ij} + z_{ijk},$$

and (1) is completely specified. Similarly, if $\sigma_s^2 = 0$,

$$(4) \quad x_{ijk} = \mu + a_i + z_{ijk},$$

and again (1) is completely specified.

We wish to test the hypothesis $H_0: \sigma_a^2 = 0$ against the alternative $H_1: \sigma_a^2 > 0$. Now let us assume we have the completely specified model given by (3). Then $\sigma_b^2 > 0$, and we obtain the analysis of variance given in Table 1. Then it follows from the likelihood ratio principle that the appropriate test procedure is to calculate the test statistic

$$(5) \quad F_0 = \frac{V_2}{V_1}$$

and to reject H_0 if $F_0 \geq F_\alpha(n_1, n_2)$, where α is the prescribed level of significance. This test is the never-pool test.

Next let us assume the completely specified model given by (4). Now the expected mean squares of Table 1 no longer include the σ_b^2 component, since $\sigma_b^2 = 0$. Application of the likelihood ratio test procedure to this model for the test of H_0 gives us the test criterion

$$(6) \quad F_0 = \frac{(n_1 + n_2)V_2}{n_1V_1 + n_2V_2}$$

and the rule to reject H_0 if $F_0 \geq F_\alpha(n_2, n_1 + n_2)$. This gives us the always-pool test.

Finally, we assume $\sigma_b^2 \geq 0$ and, hence, are confronted with the incompletely specified model given by (2). Ordinarily this model (2) might be assumed when there exists some uncertainty as to whether $\sigma_b^2 = 0$ or $\sigma_b^2 > 0$. In such cases of incomplete specification, attempts are often made to resolve the uncertainty by first testing the hypothesis that $\sigma_b^2 = 0$. The model finally chosen and, hence, the final test (test of H_0) depend upon the outcome of this original test. When the original and final tests are performed on the same set of data, the original test is referred to as a preliminary test of significance. In our example the preliminary test becomes the test of $H'_0: \sigma_b^2 = 0$ against $H'_1: \sigma_b^2 > 0$. Again, a likelihood ratio test procedure is available for this preliminary test. The statistic $F_0 = V_2/V_1$ is calculated and H'_0 rejected at the level α_1 (usually different from α) if $F_0 \geq F_{\alpha_1}(n_2, n_1)$. If H'_0 is rejected, the non-pooling test procedure indicated by (5) is used for the final test. If H'_0 is not rejected, the pooling procedure indicated by (6) is used for the final test.

It should be noted that when the final test is carried out, the model is assumed to be completely specified, that is, to be either model (3) or model (4), according as the preliminary test is found to be significant or not significant, respectively.

The essential features of the sometimes-pool procedure as applied to the component of variance model described may be summarized as follows:

(i) The three mean square V_i are independently distributed as $\chi_i^2 \sigma_i^2 / n_i$, where χ_i^2 is the (central) χ^2 statistic for n_i degrees of freedom;

(ii) The main purpose of the analysis is to test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative $\sigma_1^2 > \sigma_2^2$;

(iii) The error mean square V_2 has an expectation σ_2^2 which is greater than or equal to the expectation, σ_1^2 , of the doubtful error mean square V_1 , which may or may not be pooled.

It is clear that the above hierarchical classification is not the only analysis of variance situation giving rise to the above conditions. As an example we may quote the two-way classification with both factors random and cell repetition. Here V_1 would play the part of the within-cell mean square, while V_2 would be represented by the residual in the two-way analysis.

Aside from the preliminary test for the complete specification of the model, it is to be noted that we have made the assumptions usually made in the customary analysis of variance, namely those associated with an additive analysis of variance model. It is sometimes correctly argued that these latter assumptions may not be justified in certain situations, and in others may represent only an approximation to the actual mechanisms generating the data. This issue is, of course, one affecting the analysis of variance tests in general, and has led to extensive studies of the validity of these tests when the basic assumptions are not completely satisfied. If there is some doubt regarding the detailed assumptions for the analysis of variance model, it should be possible also to formulate the problem as being incompletely specified in these other respects. We are not concerned with these issues here. In extending the analysis of variance theory based on the assumption of linear models, our results are, strictly speaking, limited to situations in which these other assumptions are satisfied. However, the classical

TABLE 2
Mixed model example—Analysis of variance

Source of variation	df	Mean square	Exp. mean square
Between rations	$n_1 = k - 1$	V_2	$\sigma_a^2 = \sigma_a^2 + m\sigma_d^2 + mn\theta_{(a)}$
Reps \times rations	$n_2 = (k - 1)(n - 1)$	V_3	$\sigma_{\alpha}^2 = \sigma_a^2 + m\sigma_d^2$
Within pens	$n_3 = nk(m - 1)$	V_1	$\sigma_1^2 = \sigma_a^2$

analysis of variance tests have been found to be remarkably robust, that is, not sensitive to certain deviations from the basic assumptions.⁵ We expect, therefore, that our present results will likewise be applicable as useful approximations to a wider class of situations.

1.3 *Reduction of mixed models to random models.* The preceding section has been devoted to random models only. Another frequently occurring type of model is the mixed model, in which one of the factors is fixed and the other factors are random, and the hypothesis of interest is concerned with the fixed factor. A typical example of an experiment giving rise to this type of model is a randomized block experiment in which k rations are fed to each of m animals of a pen in each of n replicates. Then a suitable model for these data is given by

$$x_{tij} = \mu + a_i + b_i + d_{ti} + z_{tij},$$

where the replicate variates b_i , error variates d_{ti} , and within-pen error variates z_{tij} are assumed to be random samples from the respective normal populations $N(0, \sigma_b^2)$, $N(0, \sigma_d^2)$, and $N(0, \sigma_z^2)$, while the ration means a_i are fixed parameters. The analysis of variance based on this model is shown in Table 2. Here $\theta_{(a)} = \sum (a_i - \bar{a})^2 / (k - 1)$. Following the same general consideration of Section 1.2, it is shown in Section 2.5 how the sometimes-pool procedure for this model can be reduced to that of the random model.

1.4 *Related papers and objectives of present investigation.* The problem to be discussed here is from a general area of preliminary tests of significance. Work in this area includes studies by Bancroft [1], [2]; Mosteller [12]; Paull [14], [15]; Kitagawa [10]; Bechhofer [4]; and Bennett [5]. Paull [14], [15] studied the size and the power for the component of variance model described in Section 1.2. However, he was able to express the size and power in closed form for the case $n_3 = 2$ only, so that all comparisons made by him are restricted to that value of n_3 .

The object of the present study is to provide the necessary extension of Paull's investigation to cover all of the important degrees of freedom combinations occurring in the analyses of variance under discussion. This extension was made possible by

(i) the development of the power integrals as series formulas for even values of the degrees of freedom n_1 , n_2 , and n_3 ;

⁵ See, e.g., Box ([6], [7]) and the numerous references to earlier work given there.

TABLE 3
Component of variance model—Analysis of variance

Source of variation	Mean square	df	Exp. mean square
Treatments.....	V_3	n_3	σ_3^2
Error.....	V_2	n_2	σ_2^2
Doubtful error.....	V_1	n_1	σ_1^2

(ii) the derivation of recurrence formulas for the power for even values of n_1 , n_2 , and n_3 ;

(iii) the development of approximate formulas valid for large degrees of freedom for even values of n_1 , n_2 , and n_3 .

2. Exact and approximate formulas for power. Component of variance model.

2.1 *Mathematical formulation of the pooling procedure.* We now derive formulas for the power and size of the pooling procedure applied to the component of variance model described in Section 1. Let us first state the procedure in mathematical terms. We are given an analysis of variance as shown in Table 3.

We are interested in testing the hypothesis $H_0: \sigma_3^2 = \sigma_2^2$ against the alternative $H_1: \sigma_3^2 > \sigma_2^2$ when it is known that $\sigma_3^2 \geq \sigma_2^2 \geq \sigma_1^2$. We assume the sums of squares $n_i V_i$ are independently distributed as $\chi_i^2 \sigma_i^2$, where χ_i^2 is the central χ^2 statistic based on n_i degrees of freedom. The test procedure with sometimes pooling V_2 and V_1 is then as follows: Reject H_0 if

$$(7) \quad \begin{cases} \text{either} & \{V_2/V_1 \geq F_{n_2, n_1}(\alpha_1) \text{ and } V_3/V_2 \geq F_{n_3, n_2}(\alpha_2)\} \\ \text{or} & \{V_2/V_1 \leq F_{n_2, n_1}(\alpha_1) \text{ and } V_3/V \geq F_{n_3, n_1+n_2}(\alpha_3)\}, \end{cases}$$

where $V = (n_1 V_1 + n_2 V_2)/(n_1 + n_2)$ and $F_{n_i, n_j}(\alpha)$ is the upper 100α % point of the F -distribution with numerator df = n_i and denominator df = n_j .

The probability, P , of rejecting H_0 , which in general is the power of the test procedure, is a function of the degrees of freedom, n_1 , n_2 , and n_3 , the ratios, $\theta_{32} = \sigma_3^2/\sigma_2^2$ and $\theta_{21} = \sigma_2^2/\sigma_1^2$, and the levels of significance employed, α_1 , α_2 , and α_3 . In the special case when $\theta_{32} = 1$, this power is equal to the size of the test, i.e., the probability of type one error. In general the power P is obtained as the sum of two components corresponding to the mutually exclusive alternatives headed by either, and/or in its definition above, namely,

$$(8) \quad P_1 = \Pr \{V_2/V_1 \geq F_{n_2, n_1}(\alpha_1) \text{ and } V_3/V_2 \geq F_{n_3, n_2}(\alpha_2)\},$$

$$(9) \quad P_2 = \Pr \{V_2/V_1 \leq F_{n_2, n_1}(\alpha_1) \text{ and } V_3/V \geq F_{n_3, n_1+n_2}(\alpha_3)\}.$$

2.2 *Integral expressions for the power.* Definitive integrals for P_1 and P_2 will now be derived. The joint density of V_1 , V_2 , and V_3 is given by

$$c_1 V_1^{n_1-1} V_2^{n_2-1} V_3^{n_3-1} \exp \left\{ -\frac{1}{2} \left(\frac{n_1 V_1}{\sigma_1^2} + \frac{n_2 V_2}{\sigma_2^2} + \frac{n_3 V_3}{\sigma_3^2} \right) \right\},$$

where c_1 is a constant independent of V_1 , V_2 , and V_3 . By introducing new variates,

$$u = \frac{n_2 V_2}{\theta_{21} n_1 V_1}, \quad v = \frac{n_3 V_3}{\theta_{32} n_2 V_2}, \quad w = \frac{n_1 V_1}{n_3},$$

and integrating out w , we obtain for the joint distribution of u and v

$$f(u, v) = \frac{k u^{\frac{1}{2}(n_2+n_3)-1} v^{\frac{1}{2}n_3-1}}{(1+u+uv)^{\frac{1}{2}(n_1+n_2+n_3)}}$$

where

$$k = \frac{1}{B(n_1/2, n_2/2)B(n_3/2, (n_1+n_2)/2)}$$

The probability of rejecting the hypothesis H_0 is obtained by integrating $f(u, v)$ over the two ranges of variation of u and v which correspond to the two alternatives either and/or of definition (7). These ranges are respectively given by either

$$\frac{u_1^0}{\theta_{21}} \leq u < \infty, \quad \frac{u_2^0}{\theta_{32}} \leq v < \infty$$

or

$$0 \leq u \leq \frac{u_1^0}{\theta_{21}}, \quad \frac{u_2^0(1+\theta_{21}u)}{\theta_{32}\theta_{21}u} \leq v < \infty.$$

where

$$(10) \quad u_1^0 = \frac{n_2}{n_1} F_{n_2, n_1}(\alpha_1), \quad u_2^0 = \frac{n_3}{n_2} F_{n_3, n_2}(\alpha_2).$$

and

$$u_3^0 = \frac{n_3}{n_1 + n_2} F_{n_3, n_1+n_2}(\alpha_3).$$

Hence the formulas for the two power components become

$$P_1 = \int_a^\infty \int_d^\infty f(u, v) dv du$$

and

$$P_2 = \int_0^a \int_{c(1+\theta_{21}u)/u}^\infty f(u, v) dv du.$$

where

$$(11) \quad a = \frac{u_1^0}{\theta_{21}}, \quad c = \frac{u_2^0}{\theta_{21}\theta_{32}}, \quad d = \frac{u_2^0}{\theta_{32}}, \quad \theta_{21} \geq 1, \\ \theta_{32} \geq 1, \quad a \leq u_1^0, \quad d \leq u_2^0$$

2.3 *Exact formulas.*2.3.1 *Series formulas.*

$$P_1 = k \int_a^\infty \int_d^\infty \frac{u^{\frac{1}{2}(n_1+n_2)-1} v^{\frac{1}{2}n_2-1}}{(1+u+uv)^{\frac{1}{2}(n_1+n_2+n_3)}} dv du.$$

The transformation $z = (1+u)/(1+u+uv)$ yields

$$P_1 = k \int_a^\infty \int_0^{x_1} \frac{z^{\frac{1}{2}(n_1+n_2)-1} (1-z)^{\frac{1}{2}n_2-1} u^{\frac{1}{2}n_2-1}}{(1+u)^{\frac{1}{2}(n_1+n_2)}} dz du,$$

where

$$x_1 = \frac{1+u}{1+u(1+d)}.$$

The binomial expansion of $(1-z)^{\frac{1}{2}(n_2)-1}$ gives us

$$P_1 = k \int_a^\infty \int_0^{x_1} \frac{u^{\frac{1}{2}n_2-1} f(z)}{(1+u)^{\frac{1}{2}(n_1+n_2)}} dz du,$$

where

$$f(z) = \sum_{j=0}^{\frac{1}{2}n_2-1} (-1)^j \binom{n_2/2-1}{j} z^{\frac{1}{2}(n_1+n_2)+j-1}$$

Upon performing the integrations with respect to u and z , we obtain

$$P_1 = \sum_j \left[\frac{(-1)^{j-1} \binom{n_2/2-1}{j-1}}{[(n_1+n_2)/2+j-1] B(n_1/2, n_2/2) B(n_2/2, (n_1+n_2)/2) (1+d)^{n_2/2}} \right] \\ \times \left[\sum_{r=0}^{j-1} \frac{\binom{j-1}{r} B(n_1/2+j-1-r, n_2/2+r)}{(1+d)^r} \cdot I_{x_2}(n_1/2+j-1-r, n_2/2+r) \right],$$

where

$$(12) \quad x_2 = \frac{1}{1+a+ad}.$$

We now consider

$$P_2 = k \int_0^a \int_{x_3}^\infty \frac{u^{(n_1+n_2)/2-1} v^{n_2/2-1}}{(1+u+uv)^{(n_1+n_2+n_3)/2}} dv du,$$

where

$$x_3 = [c(1+\theta_{21}u)]/u = (c+bu)/u,$$

and

$$(13) \quad b = \frac{u_{21}^0}{\theta_{22}}.$$

Using procedures similar to those used in deriving P_1 , we obtain

$$P_2 = \sum_j \left\{ \frac{(-1)^{j-1} \binom{n_2/2 - 1}{j-1}}{((n_1 + n_2)/2 + j - 1) B(n_1/2, n_2/2) \cdot B(n_2/2, (n_1 + n_2)/2) (1+b)^{n_1/2} (1+c)^{n_1/2}} \right. \\ \left. \times \left[\sum_{r=0}^{j-1} \frac{\binom{j-1}{r} B[(n_2/2) + r, (n_1/2) + j - 1 - r] \cdot I_{x_4}[(n_2/2) + r, (n_1/2) + j - 1 - r]}{(1+b)^r (1+c)^{j-1-r}} \right] \right\},$$

where

$$(14) \quad x_4 = \frac{a(1+b)}{1+c+a(1+b)}.$$

2.3.2 Recurrence formulas. Integrating P_1 (as originally given in section 2.2) by parts with respect to v , we obtain

$$P_1(n_3) = \frac{(d)^{n_3/2-1}}{(n_1 + n_2 + n_3)/2 - 1} k \int_a^\infty \frac{u^{(n_2+n_3)/2-2}}{(1+u(1+d))^{(n_1+n_2+n_3)/2-1}} du \\ + \frac{(n_3/2 - 1)k}{(n_1 + n_2 + n_3)/2 - 1} P_1(n_3 - 2).$$

Upon integrating with respect to u , we obtain

$$(15) \quad P_1(n_3) = \frac{(d)^{n_3/2-1} I_{x_2}(n_1/2, (n_2 + n_3)/2 - 1)}{(n_3/2 - 1) B(n_3/2 - 1, n_2/2) (1+d)^{(n_1+n_3)/2-1}} + P_1(n_3 - 2),$$

where x_2 is given by (12). For the set of initial values at $n_3 = 2$ it is found that

$$(16) \quad P_1(2) = \frac{I_{x_2}(n_1/2, n_2/2)}{(1+d)^{n_2/2}}.$$

The recurrence development for P_2 is similar but more cumbersome. We obtain the relation

$$(17) \quad P_2(n_1, n_3) = \frac{1}{1+c} \frac{(1/a)^{n_1/2-1} I_{x_3}((n_1 + n_2)/2 - 1, n_3/2)}{((n_1 + n_2)/2 - 1) B(n_1/2, n_2/2) (1+1/a)^{(n_1+n_2)/2-1}} \\ + c \cdot P_2(n_1, n_3 - 2) + P_2(n_1 - 2, n_3),$$

where

$$(18) \quad x_3 = \frac{1 + (1/a)}{1 + (1/a) + b + (c/a)}.$$

The formulas for the initial values are found to be

$$(19) \quad P_2(n_1, 2) = \frac{I_{x_4}[(n_2/2), (n_1/2)]}{(1+b)^{n_2/2} (1+c)^{n_1/2}}$$

and

$$(20) \quad P_2(2, n_3) = \frac{1}{1+c} \left\{ \frac{I_{x_2}[(n_2/2), (n_3/2)]}{[1 + (1/a)]^{n_2/2}} + c \cdot P_2(2, n_3 - 2) \right\},$$

where x_2 and x_4 are given by (12) and (14), respectively.

2.4 Approximate formulas. We now derive simpler approximate formulas. We first consider P_2 . Writing $F_1 = F_{n_2, n_1}(\alpha_1)$, $F_2 = F_{n_2, n_2}(\alpha_2)$, $F_3 = F_{n_1, n_1+n_2}(\alpha_3)$, we have

$$P_2 = \Pr \{V_2/V_1 \leq F_1 \text{ and } V_3/V \geq F_3\}.$$

As $n_1 \rightarrow \infty$ both $V_1 \rightarrow \sigma_1^2$ and $V \rightarrow \sigma_1^2$ and, in the limit, the two ratios V_2/V_1 and V_3/V are independently distributed. It is therefore suggested that for large n_1 we use the approximation

$$(21) \quad \begin{aligned} P_2 &\doteq \Pr \{V_2/V_1 \leq F_1\} \Pr \{V_3/V \geq F_3\} \\ &\doteq [1 - I_{x_6}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2)] I_{x_7}(\tfrac{1}{2}(n_1 + n_2), \tfrac{1}{2}n_3), \end{aligned}$$

where

$$\begin{aligned} x_6 &= 1 / \left(1 + \frac{1 - x(\alpha_1)}{\theta_{21} x(\alpha_1)} \right), \\ x_7 &= (n_1 + n_2) / \left(n_1 + n_2 + \frac{(n_2 \theta_{21} + n_1)(1 - x(\alpha_3))}{\theta_{21} \theta_{32} x(\alpha_3)} \right), \end{aligned}$$

and $x(\alpha_1)$, $x(\alpha_3)$ are respectively the roots, x , of $I_x(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) = \alpha_1$ and $I_x(\tfrac{1}{2}(n_1 + n_2), \tfrac{1}{2}n_3) = \alpha_3$. Here we have used the well-known relation between the incomplete Beta function $I_x(a, b)$ and the F -integral, viz,

$$\Pr \{F_{v_1, v_2} \leq F_0\} = I_x(\tfrac{1}{2}v_1, \tfrac{1}{2}v_2), \text{ with } x = v_1 F_0 / (v_2 + v_1 F_0).$$

We have also used the approximation that for large n_1 , V is approximately distributed as $(n_1 \sigma_1^2 + n_2 \sigma_2^2) \chi_{n_1+n_2}^2 / (n_1 + n_2)^2$.

Next we turn to

$$P_1 = \Pr \{V_2/V_1 \geq F_1 \text{ and } V_3/V_2 \geq F_2\}.$$

Here we could use a similar argument if we were to let $n_2 \rightarrow \infty$. This limit would however, not yield useful results. The important situation in pooling procedures, is one in which n_2 is moderate or small. Instead we use the well-known normal approximation to $\log V_i$. M. S. Bartlett and D. G. Kendall [3] have shown that $\log V_i$ is approximately $N(\log \sigma_i^2, 2/(n_i - 1))$, provided that n_i is not too small. Writing

$$u = \log V_2 - \log V_1 \text{ and } z = \log V_3 - \log V_2,$$

it follows that the joint distribution of u and z is approximately bivariate normal with correlation coefficient

$$(22) \quad \rho = -1 / \left\{ \left(1 + \frac{n_2 - 1}{n_3 - 1} \right) \left(1 + \frac{n_2 - 1}{n_1 - 1} \right) \right\}^{\frac{1}{2}}.$$

We may therefore employ the tables of the double probability integral of a bivariate normal surface of K. Pearson [16]. Tables VIII and IX. If x and y follow a bivariate normal distribution with both means equal to 0, correlation coefficient ρ , and both standard deviations equal to unity, then these tables give the probabilities $P_T(h, k)$, for $x \geq h$ and $y \geq k$. In our case, ρ is given by (22) and h and k by

$$h = \frac{2z_{n_1, n_1}(\alpha_1) - \log \theta_{21}}{[2/(n_1 - 1) + 2/(n_2 - 1)]^{1/2}}, \quad k = \frac{2z_{n_2, n_2}(\alpha_2) - \log \theta_{22}}{[2/(n_2 - 1) + 2/(n_3 - 1)]^{1/2}},$$

where $z_{n_i, n_j}(\alpha)$ is the upper 100α per cent point of Fisher's z distribution with numerator degrees of freedom n_i and denominator degrees of freedom n_j .⁶

2.5 *Theory of reduction of mixed model to random model.* Certain mixed models of analysis of variance were described in Section 1.3. No new formulas are required for these models, as we shall show that the joint distribution of the three mean squares is, at least approximately, equal to that of the component of variance model. The exact specifications of the distribution for the mixed model being considered are as follows. (Primed parameters will be used to specify the parameters for the mixed model.)

(a) The error mean square V_2 and the doubtful error mean square V_1 are distributed as $\chi^2_i \sigma_i^2 / n'_i$ ($i = 1, 2$), where χ^2_i is the central χ^2 statistic with n'_i degrees of freedom. On the other hand, the treatment mean square V_3 is distributed as $\chi'^2_{n_3} \sigma_3^2 / n'_3$, where $\chi'^2_{n_3}$ is the noncentral χ^2 statistic with n'_3 degrees of freedom and noncentrality parameter

$$\lambda = \frac{n'_3 \sigma_3^2 - n'_3 \sigma_2^2}{2\sigma_2^2} = \frac{n'_3}{2} (\theta'_{32} - 1),$$

where $\theta'_{32} = \sigma_3^2 / \sigma_2^2$. V_1 , V_2 , and V_3 are independent.

(b) The main purpose of the analysis is to test the hypothesis $H_0: \sigma_3^2 = \sigma_2^2$ against the alternative $H_1: \sigma_3^2 > \sigma_2^2$.

(c) The true error mean square, V_2 , has an expectation σ_2^2 which is greater than or equal to the expectation, σ_1^2 , of the doubtful error mean square.

The probability P of rejecting H_0 is obtained as the sum of the two components,

$$(23) \quad P_1 = \Pr \{V_2/V_1 \geq F'_{n_2, n'_1}(\alpha'_1) \text{ and } V_3/V_2 \geq F'_{n_3, n'_2}(\alpha'_2)\}$$

and

$$(24) \quad P_2 = \Pr \{V_2/V_1 < F'_{n_2, n'_1}(\alpha'_1) \text{ and } V_3/V \geq F'_{n_3, n'_1+n'_2}(\alpha'_3)\}.$$

In evaluating these probabilities we use, the approximation first used by Patnaik [13]. We replace $\chi'^2_{n_3}$ by $C\chi^2_{\nu_3}$, $\chi^2_{\nu_3}$ being the central χ^2 statistic based upon ν_3 degrees of freedom, where

$$\nu_3 = n'_3 + \frac{4\lambda^2}{n'_3 + 4\lambda} \text{ df}$$

⁶ See Appendix Table 4 for illustrations of the nature of the approximation to the integral P_1 .

TABLE 4

Modified parameters for random model corresponding to specified parameters for mixed model

Specified parameters for mixed model	Modified parameters for random model
n_1'	$n_1 = n_1'$
n_2'	$n_2 = n_2'$
n_3'	$n_3 = \nu_3 = n_3' + \frac{4\lambda^2}{n_3' + 4\lambda}$
α_1'	$\alpha_1 = \alpha_1'$
α_2'	$\alpha_2 = \text{Root of } F_{n_3', n_2'}(\alpha_2') = F_{\nu_3, n_2'}(\alpha_2')$
α_3'	$\alpha_3 = \text{Root of } F_{n_3, n_1+n_2}(\alpha_3') = F_{\nu_3, n_1+n_2}(\alpha_3')$
$\theta_{21}' = \sigma_2^2/\sigma_1^2$	$\theta_{21} = \theta_{21}'$
$\theta_{32}' = \sigma_3^2/\sigma_2^2$	$\theta_{32} = (2\lambda + n_3')/n_3'$

and

$$C = 1 + \frac{2\lambda}{n_3' + 2\lambda}.$$

Since the use of this approximation reduces the noncentral χ^2 statistic to a central χ^2 statistic, all three statistics are now central.

We now compare the power for the mixed model as defined by (23) and (24) with the corresponding formulas (8) and (9), for modified values of the eight parameters as indicated in Table 4. Entering the random model tables with these altered parameters we obtain the mixed model power. It will be seen that when we deal with the size for the mixed model we have $\lambda = 0$ and hence $\nu_3 = n_3'$, so that all primed parameters agree with those without primes. Thus our entire size discussion to follow is directly applicable to the mixed model. On the other hand, the power evaluations, which refer to $\alpha_2 = \alpha_3 = .05$, will in general provide answers for larger values of α_2' and α_3' , and these levels α_2' and α_3' will vary with λ . For a proper evaluation of power corresponding to a given pair of significance levels α_2' and α_3' , say, $\alpha_2' = \alpha_3' = .05$, a more extensive tabulation of (8) and (9) as described in the subsequent section, would be required.

2.6 Application of derived formulas. The recurrence formulas derived in Section 2.3.2 were used to construct master tables of P_1 and P_2 . These master tables were constructed for

$$\frac{n_2}{2} = 5, \quad \frac{n_1}{2} = 3(1) 10, \quad \text{and} \quad \frac{n_3}{2} = 1(1) 6.$$

Also, tables were constructed for $\frac{1}{2}n_2 = 3$, in order that the effect of small error degrees of freedom could be better studied. However, the latter tables were confined to the values

$$\frac{1}{2}n_2 = 1 \quad \text{and} \quad \frac{1}{2}n_1 = 4, 7, \quad \text{and} \quad 10.$$

To compute the power component P_1 in a given problem from these master tables, for specified degrees of freedom n_1 , n_2 , and n_3 and levels of significance α_1 , α_2 , the values of the parameters u_1^0 and u_2^0 are computed from (10). From these values and those specified for θ_{21} and θ_{22} , the corresponding values of a and d from (11) and hence the value of x_2 from (12) are computed. P_1 is then obtained by interpolation in the appropriate master table.

The procedure used to compute the component P_2 was similar, and required evaluation of an interpolation for the parameters a , b , c , x_1 and x_2 ; but interpolation with respect to a was avoided by choosing values of θ_{21} which would result in tabular values of a . This accounts for the decimal values of θ_{21} found in our tables.

The approximate formula (21) for P_2 derived in Section 2.5 is exact for $n_1 = \infty$ and was found to be very effective for large n_1 , yielding either P_1 values directly to sufficient accuracy, or facilitating extrapolation of the master tables.

3. Discussion of power and size curves and comparison of test procedures.

3.1 *Type of recommendations attempted.* We have seen that the power of our test procedures depends upon the following eight parameters: the degrees of freedom n_1 , n_2 , and n_3 ; the variance ratios $\theta_{21} = \sigma_2^2/\sigma_1^2$, $\theta_{22} = \sigma_3^2/\sigma_2^2$; and the levels of significance α_1 , α_2 , α_3 . Of these, the degrees of freedom n_1 , n_2 , and n_3 are completely determined by the analysis of variance table, while the variance ratios are generally unknown (except in the case of the size of the procedure, when $\theta_{22} = 1$). Any recommendations that are to be made must therefore be confined to the levels of significance, α_1 , α_2 , and α_3 . We shall here be primarily concerned with the size of the procedure being in the vicinity of .05. It will be apparent from what follows that a convenient way to achieve this is to choose $\alpha_2 = \alpha_3 = .05$, that is, to choose procedures in which the significance levels of both final tests are .05. However, the remaining parameter, α_1 , the level of significance for the preliminary test, is entirely at our disposal. In attempting recommendations, therefore, we shall be concerned with the choice of the level of α_1 . Should α_1 be, say, .05, .25, .50, or should we use what Paull ([14], p. 4; [15]) has called the borderline test, where α_1 will be near .70 to .80? In choosing the level of α_1 , we shall consider

- (i) the variation in the size of our test procedure as a function of the parameter θ_{21} , and
- (ii) a comparison of the power of our test procedure with that of the never-pool test of the same size.

3.2 *Size.* The size of our test procedure does not equal the nominal level of .05, but varies about this level as α_1 and θ_{21} vary. Figures 1 to 10⁷ give us examples of size curves, illustrating the variations in type one error with variation in θ_{21} for fixed values of the remaining parameters.

⁷ A selection of figures and tables has been assembled in the Appendix. Additional size and power curves and tables, illustrating the points to be made in the ensuing discussion will be found in [8].

Note that as θ_{21} becomes large, the size approaches .05; for, as $\theta_{21} \rightarrow \infty$, the preliminary test will almost certainly be significant, pooling will almost certainly not occur, and hence the final test will almost certainly be that of V_3/V_2 , having a size of .05.

At the lower extreme, that is, at $\theta_{21} = 1$, the size is at its minimum, which is less than .05. This minimum, and even more so the size peak, are points of particular interest.

We first consider the size peak. Referring to the size curves for a preliminary test carried out at the 5% level (see Figs. 1 and 2), we note that the peak is usually very high. Clearly, a preliminary test carried out at this level will in many cases admit an unacceptable size disturbance. This is due to the fact that at this level, the preliminary test will frequently admit pooling V_2 and V_1 when σ_1^2 is smaller than the true error mean square σ_2^2 , and thereby increase the probability of type one error. We therefore seek a preliminary test in which pooling is admitted less readily; we next investigate the level $\alpha_1 = .25$. At this level (compare Figs. 2 and 3), size control is considerably better, and in many

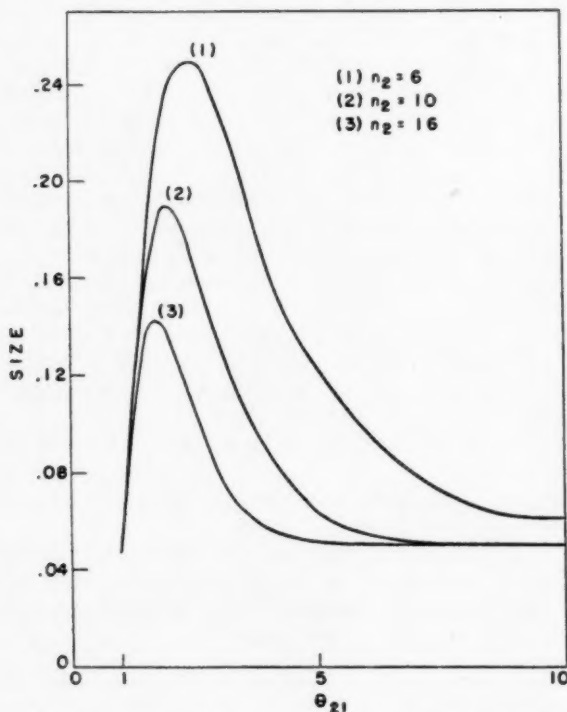


FIG. 1. Size curves for $n_1 = \infty$, $n_2 = 6$, $\alpha_1 = \alpha_2 = \alpha_3 = .05$

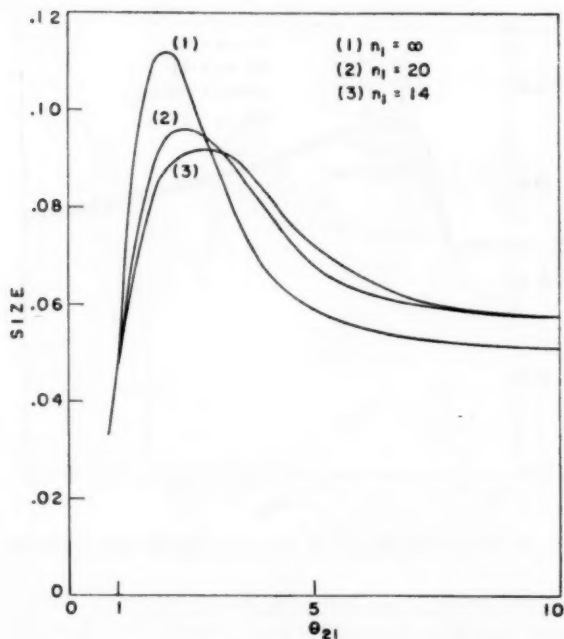


FIG. 2. Size curves for $n_3 = 2$, $n_2 = 10$, $\alpha_1 = \alpha_2 = \alpha_3 = .05$

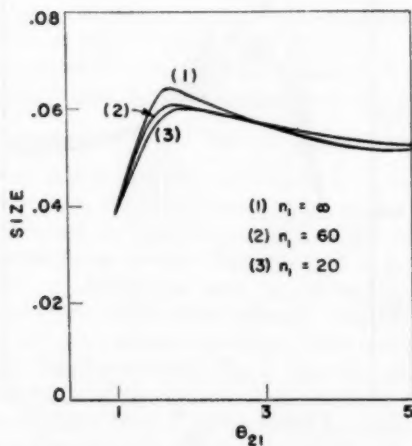
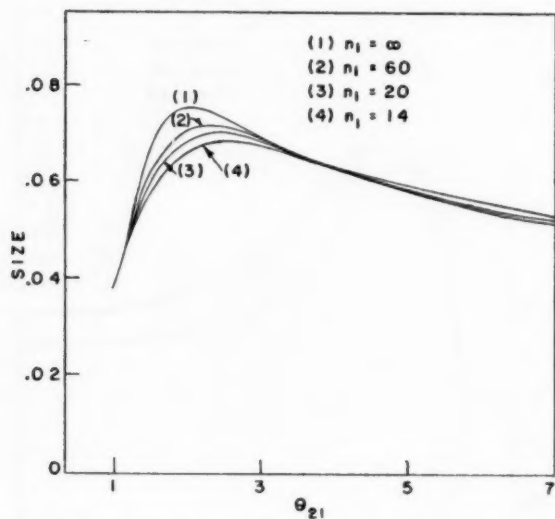
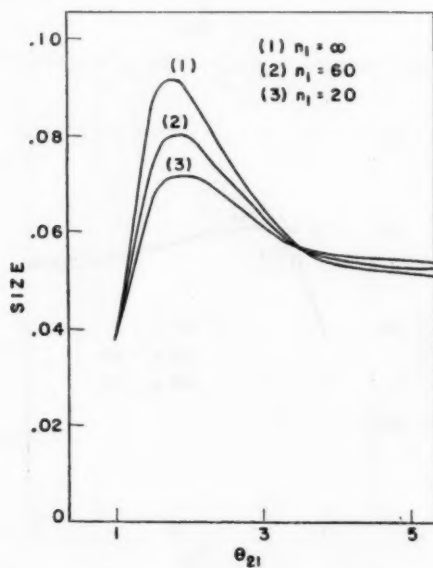


FIG. 3. Size curves for $n_3 = 2$, $n_2 = 10$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

FIG. 4. Size curves for $n_2 = 2$, $n_2 = 6$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$ FIG. 5. Size curves for $n_2 = 6$, $n_2 = 10$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

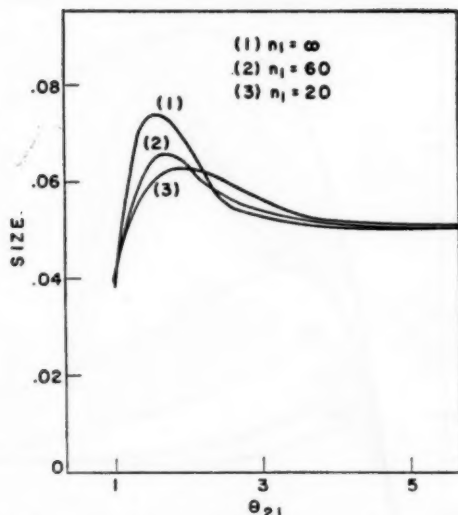


FIG. 6. Size curves for $n_3 = 6$, $n_2 = 16$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

cases the peaks do not go beyond .08. (See, e.g., Figs. 3, 4, 5, and 6.) It is observed that, in general, the size peak increases as n_1 or n_3 increases or as n_2 decreases. (See Figs. 1 through 6.)

It is of course quite arbitrary to specify any rules for maintaining an acceptable upper tolerance for the size peak, since what is considered acceptable is a matter of opinion. In using a nominal size of .05, if we stipulate that our size peak should not go much beyond 10 per cent, then we find that even with the 25 per cent preliminary level, there are situations in which this upper limit is exceeded. Generally speaking, these unacceptable size peaks occur when

$$(25) \quad n_3 \geq n_2 \quad \text{and} \quad n_1 \geq 5n_2.$$

(It should be noted that the occurrence of $n_3 > n_2$ is clearly rare.) This means that when the treatment degrees of freedom are greater than or equal to the error degrees of freedom, we must be careful if at the same time the doubtful error degrees of freedom are greater than or equal to five times the true error degrees of freedom; or, briefly, we must be careful when pooling promises a large gain in the precision of the error estimate. This rule has been established by an empirical study of an extensive number of size curves, and is not based on any analytic study. See, for example, Fig. 7. Here the situation represented by Curve 1 would be excluded by our rule. See also Figs. 8 and 9, in which the situations represented by Curves 1 and 2 would be excluded. If our rule is followed, size disturbances such as are represented in Figs. 3, 4, 5, and 6 occur; and also disturbances such as are represented in Figs. 7, 8, and 9 occur, with the

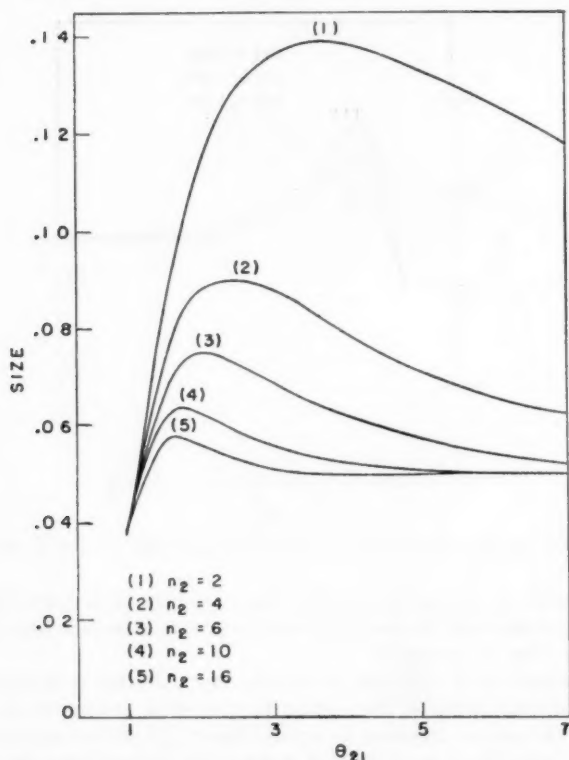


FIG. 7. Size curves for $n_1 = 2$, $n_2 = \infty$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

exception of the excluded cases mentioned above. In the situations represented by (25), a more conservative level of α_1 would be appropriate. From a study of a number of size curves it appears that a preliminary test at the 50 per cent level will ensure adequate control of the size peak in these cases. (See Fig. 9.)

Not only the size peak, but also the size minimum is affected by the level of the preliminary test. From theorems proved by Paull ([14], Chap. 4; [15]), we know that the size of our test procedures is a minimum with respect to θ_{21} at θ_{21} equal to one, and that a lower bound for the size for this value of θ_{21} is

$$(1 - \alpha_1)(.05).$$

These lower bounds are .0475, .0375, and .025 for $\alpha_1 = .05$, .25, and .50, respectively. For some of our curves the plotted minimum sizes are situated very close to these lower bounds. For the borderline test, where, as proved by Paull ([14], [15]) the size is always less than .05, this lower bound varies in magnitude from approximately .01 to approximately .015. We have computed actual

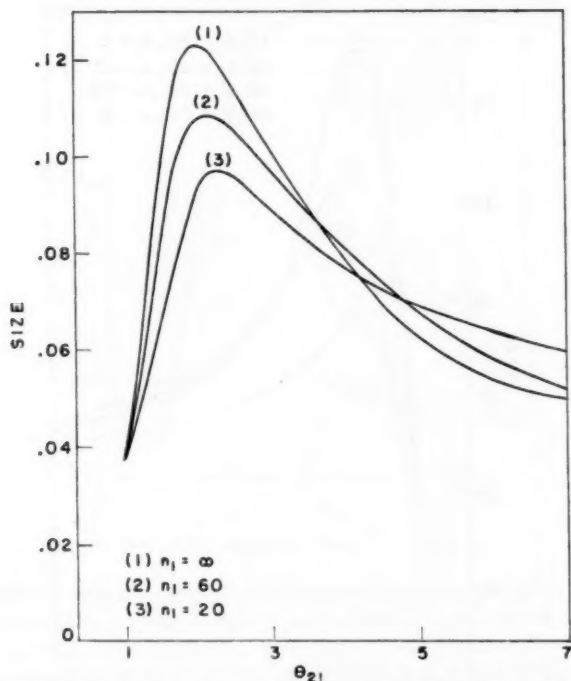
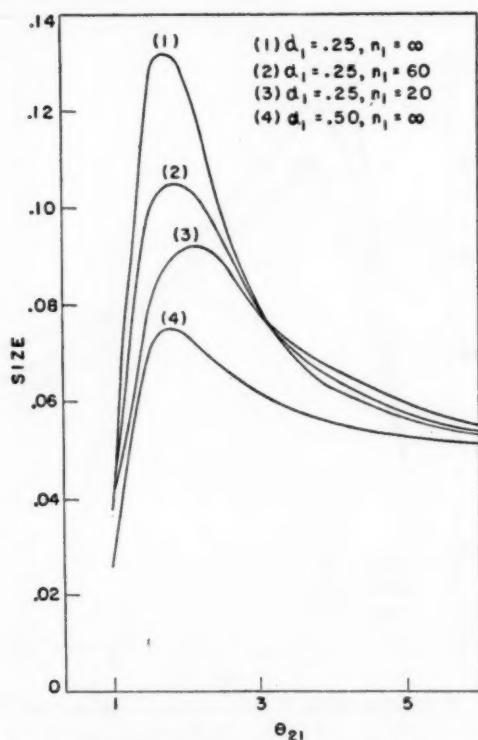


FIG. 8. Size curves for $n_1 = 6$, $n_2 = 6$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

minimum size values for the borderline test for selected values of n_1 , n_2 , and n_3 . For small n_2 and n_3 , these are very close to their lower bounds, irrespective of n_1 . A person using this test should therefore remember that he may be using a test which has a considerably lower size than .05. The actual disturbance is of course small, but the proportional disturbance is considerable. However, since the borderline test size disturbance is a reduction rather than an increase in size and is therefore on the conservative side, we are not attempting to make any definite rules as to when the experimenter should avoid the use of this test, but merely to remind him that large proportional size disturbances occur when n_2 and n_3 are both small (≤ 6).

Summarizing our considerations of size control, therefore, we have narrowed down our recommendable range of α_1 to $\alpha_1 \geq .25$, with the reservations that in certain cases characterized by inequalities (25), $\alpha_1 = .25$ would not be desirable, as it would admit too large a peak in the size curve; and that for very small values of n_2 and n_3 the experimenter may not wish to use the borderline test, as this would admit too low a size minimum.

The discussion thus far has been concerned with test procedures in which

FIG. 9. Size curves for $n_1 = 12$, $n_2 = 10$, $\alpha_2 = \alpha_3 = .05$

$\alpha_2 = \alpha_3 = .05$. A few special cases for $\alpha_2 = \alpha_3 = .01$ and $\alpha_1 = .25$ have also been investigated. In all these situations, larger proportional size disturbances than those found for $\alpha_2 = \alpha_3 = .05$ were experienced, even for cases which our rule would accept. (See Fig. 10.)

3.3 Frequency of pooling. We have been discussing the effect of increasing α_1 in order to achieve size control. It is obvious that for $\alpha_1 = 1$, our preliminary F per cent point would be zero, and pooling would never occur. The question arises as to the relative frequency of pooling for the intermediate values of α_1 that we have been considering. When $\theta_{21} = 1$, the probability that V_2/V_1 exceeds $F_{\alpha_1}(n_2, n_1)$ is α_1 , so that pooling occurs with relative frequency $1 - \alpha_1$. As θ_{21} increases, this frequency rapidly decreases, approaching the limit zero as θ_{21} becomes infinite. Evaluations of these frequencies of pooling show that, while for $\alpha_1 = .25$ and small values of θ_{21} , pooling will occur in the majority of experiments, when $\alpha_1 = .50$ the frequency is usually well below .50. This frequency of pooling is of course even smaller for the borderline test, where α_1

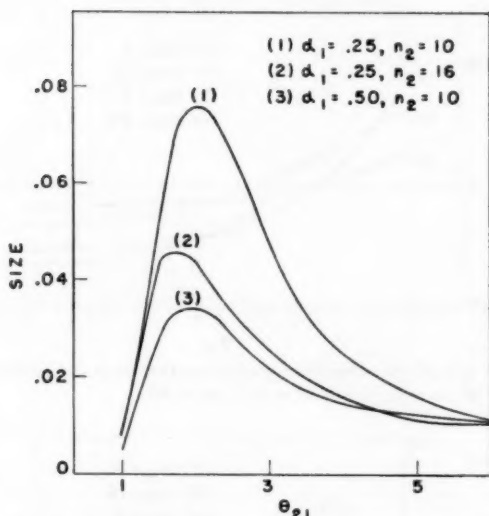


FIG. 10. Size curves for $n_2 = 12$, $n_1 = \infty$, $\alpha_2 = \alpha_3 = .01$

usually takes on values in the neighborhood of .7 to .8; when such large values of α_1 are employed, pooling occurs in only about 25 per cent of all situations for which $\theta_{21} = 1$, and this pooling percentage rapidly decreases as θ_{21} increases. While this property by itself cannot be regarded as a disadvantage of the borderline test, it is clear that, if this test were the only one recommended to the experimenter, he would hardly ever pool.

3.4 Power. We now attempt a comparison of the power of our sometimes-pool procedure with that of the never-pool test. As is well known, any comparison of power of any two test procedures is a fair comparison if the two test procedures have the same size. We have seen that the size of our sometimes-pool procedures is not at the constant level of .05, but varies about this, depending upon the parameter θ_{21} . The method of power comparison we have therefore adopted is as follows:

- (i) Assume a fixed value of the parameter θ_{21} .
- (ii) For this value of θ_{21} , evaluate the size of the sometimes-pool test.
- (iii) For this level of size, evaluate the power curve of the never-pool test; this power is then directly comparable with that of the sometimes-pool test corresponding to the chosen value of θ_{21} .

For an illustration of such comparisons, see Table 1. Here we have $n_1 = 20$, $n_2 = 6$, $n_3 = 2$, $\alpha_1 = .25$ and $\alpha_2 = \alpha_3 = .05$. For $\theta_{21} = 1$, the sometimes-pool procedure is always more powerful than the never-pool test of the same size. For $\theta_{21} = 1.5$, the powers are very similar; on the other hand, for $\theta_{21} = 2$, the never-pool test is always more powerful. See also Tables 2 and 3, which again

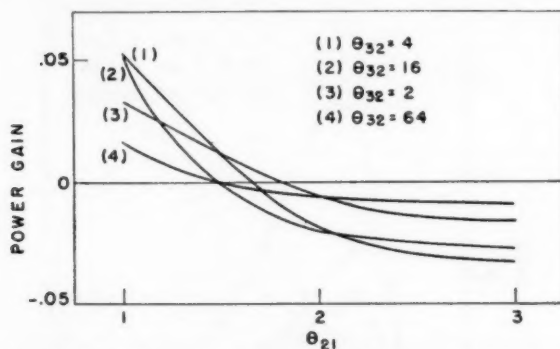


FIG. 11. Power gain of the sometimes-pool procedure over the never-pool test of the same size for $n_1 = 20$, $n_2 = 2$, $n_3 = 6$, $\alpha_1 = \alpha_2 = \alpha_3 = .05$.

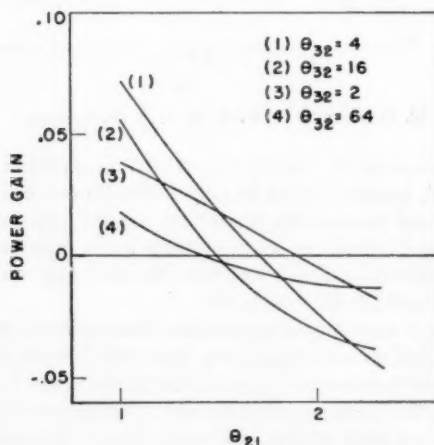


FIG. 12. Power gain of the sometimes-pool procedure over the never-pool test of the same size for $n_1 = 20$, $n_2 = 2$, $n_3 = 6$, $F_1 = 2 F_{.10}$, $\alpha_2 = \alpha_3 = .05$.

illustrate the fact that the sometimes-pool procedure is more powerful for small θ_{21} but less powerful for large θ_{21} .

In order to show more clearly the dependence of these power differences on θ_{21} , we have plotted in Figs. 11, 12 and 13 the difference between two corresponding power points against θ_{21} . Here each curve corresponds to a fixed value of θ_{32} —i.e., that value of θ_{32} at which the difference between the power ordinates of the power curves was taken. It will be seen, again, that for small θ_{21} the differences are positive (the sometimes-pool procedure is more powerful than the never-pool test), while for larger θ_{21} the position is reversed. As $\theta_{21} \rightarrow \infty$, the difference tends to 0, since both procedures tend to the never-pool test at the

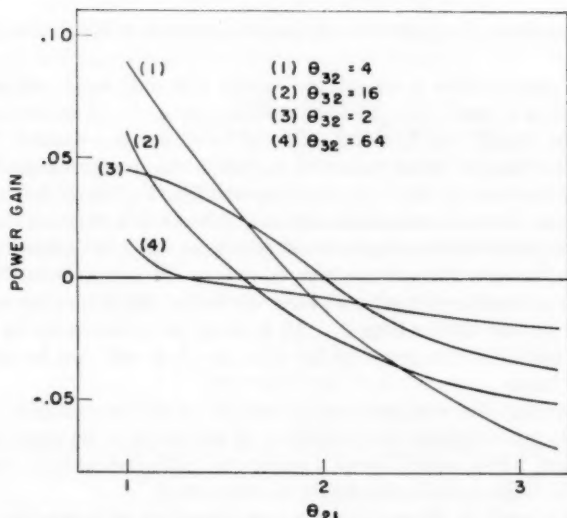


FIG. 13. Power gain of the sometimes-pool procedure over the never-pool test of the same size for $n_1 = 20$, $n_2 = 2$, $n_3 = 6$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$.

.05 level of significance. The transition from favorable to unfavorable power conditions generally occurs between $\theta_{21} = 1.5$ and $\theta_{21} = 2.0$. From other similar curves not shown here, it is seen that the magnitude of these power gains and losses increases with increasing n_3 , or decreasing n_2 , or increasing n_1 .

Figures 11, 12, and 13 also illustrate the effect of decreasing the per cent point of F for the preliminary test. In these figures, corresponding power comparisons are given, respectively, for $\alpha_1 = .05$, $F_1 = 2F_{.50}(n_2, n_1)$,^{*} and $\alpha_1 = .25$. There is a general tendency for both power gains and power losses to diminish as the per cent point of F decreases—i.e., as α_1 increases from values such as .05 through intermediate values such as .25 to the level of the borderline test (approximately $\alpha = .70$ to .80). Here the gain in power has diminished further, but the power losses have completely disappeared. In fact, a theorem by Paull ([14], p. 61; [15]) proves that the borderline test is always more powerful than the corresponding never-pool test of the same size, although the power gain is small for large θ_{21} . However, as stated in Section 3.2, this size is below the nominal level of .05. If we compare the borderline test power with that of the never-pool test at the nominal level of .05, the former is always less powerful. It is likewise less powerful than our sometimes-pool procedures for $\alpha_1 = .50$ and $\alpha_1 = .25$, which have, of course, a larger size.

We now attempt recommendations, considering the relative merits of the procedures at $\alpha_1 = .25$, $\alpha_1 = .50$, and $\alpha_1 = .7$ to .8 (the borderline level). These recommendations are somewhat subjective, since they are contingent upon what

* This means we use a preliminary test $V_2/V_1 \geq 2 F_{.50}(n_2, n_1)$.

the experimenter may regard as a reasonable assumption concerning the parameter θ_{21} .

(i) If the experimenter is reasonably certain that only small values of θ_{21} can be envisaged as a possibility, he is advised to use $\alpha_1 = .25$ except in the cases (25) when he should use $\alpha_1 = .50$, in order to ensure size control. Our figures show that the range of small values of θ_{21} , when the sometimes-pool procedure gives a gain in power, is approximately between 1 and 1.5 to 2. An experimenter about to adopt this recommendation but not quite certain about his assumptions may wish to know the consequences which result from his adopting this procedure when, in fact, unknown to him, θ_{21} is large. It is seen from Figs. 1 to 10 that in such a situation he will still have control of the size of his test; in fact the size will be near .05 for large θ_{21} . All he loses (as is illustrated by our power figures and tables) is the power of his test; this is a risk that he may well be prepared to take.

(ii) If, however, the experimenter can make no such assumption about θ_{21} , and wishes to guard against the possibility of power losses, he may then use the borderline test, which would ensure a power gain, although he must realize

(a) that for large θ_{21} this gain would be very small;

(b) that for small θ_{21} he would use a test procedure of a very much smaller size than $\alpha_1 = .05$ (particularly when n_2 and n_3 are ≤ 6) and accordingly a test which is much less powerful than the never-pool test of size .05. In fact, he may in these circumstances prefer not to pool at all.

It may be correctly argued that, in order to control the size peak, to advocate $\alpha_1 = .50$ in the cases characterized by (25), and $\alpha_1 = .25$ otherwise, introduces an artificial discontinuity in our recommendations. It would be quite feasible (although it would require a considerable effort in computation) to evaluate for any given triplet n_1, n_2 , and n_3 that value of α_1 which results in a size peak of 0.10 exactly. Since this level of α_1 would depend on the degrees of freedom n_1, n_2 , and n_3 , it would be necessary to evaluate the associated per cent points of F . For such recommendations to be useful, this table of $F_{\alpha_1}(n_1, n_2)$ (which would be a large 3 parametric table with n_1, n_2 , and n_3 as arguments) would have to be published. To encumber the experimenter with special tables for the preliminary F -test in addition to the standard F -tables for the final F -tests appeared to us to be unnecessary, and the use of the published Merrington and Thompson [11] 25% and 50% points of F preferable.

We should note here that a rule favored by Paull ([14], Chap. 6; [15]) advocating testing the ratio V_2/V_1 against $2F_{.50}(n_2, n_1)$ will not ensure adequate control of the size peak, since $2F_{.50} > F_{.25}$ in general, and we have just seen that $F_{.25}$ is sometimes too large and hence not always acceptable as a significance level for the preliminary test. Also, it would appear to us that no rule of the form $V_2/V_1 > \text{constant}$ is very satisfactory, for with such a rule the frequency with which pooling occurs, as well as the size, varies considerably with the degrees of freedom n_1 and n_2 .

Concerning recommendation ii, the experimenter would require knowledge of the precise level of α_1 for the borderline test, or, better still, the value of F as-

sociated with it. Paull ([14], p. 20; [15]) gives a simple formula from which the following is derived: F point for borderline test equals

$$\frac{(n_1 F_{n_1, n_1+n_2}(\alpha_2))}{(n_1 + n_2)(F_{n_1, n_2}(\alpha_2) - n_2 F_{n_2, n_1+n_2}(\alpha_3))},$$

where $F_{n_1, n_2}(\alpha_2)$ represents the 100 α_2 per cent point of F with numerator df n_1 and denominator df n_2 . Similar statements can be made for the other symbols.

It has been noted that the above recommendations depend upon some a priori information regarding θ_{21} . It is shown in a number of examples discussed in the Wright-Patterson report how this information can often be obtained from the general conditions under which the experiments were carried out.

APPENDIX

FIGURES AND TABLES

TABLE 1

The power of the sometimes-pool procedure and the never-pool test of the same size, for $n_1 = 20$, $n_2 = 6$, $n_3 = 2$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

θ_{21}	Test	θ_{21}				
		1	2	4	16	64
1	s.p.	.038	.161	.368	.757	.930
	n.p.	.038	.127	.314	.705	.913
1.5	s.p.	.060	.189	.385	.756	.930
	n.p.	.060	.178	.373	.757	.930
2	s.p.	.068	.190	.377	.751	.935
	n.p.	.068	.195	.396	.771	.935
3	s.p.	.068	.178	.361	.743	.926
	n.p.	.068	.194	.394	.770	.935
5	s.p.	.058	.164	.348	.738	.924
	n.p.	.058	.175	.369	.754	.930
	n.p.	.050	.156	.343	.737	.924

TABLE 2

The power of the sometimes-pool procedure and the never-pool test of the same size, for $n_1 = 20$, $n_2 = 10$, $n_3 = 12$, $\alpha = .25$, $\alpha_2 = \alpha_3 = .05$

θ_{21}	Test	θ_{21}				
		1	2	4	10	50
1.066	s.p.	.044	.354	.741	.979	1.000
	n.p.	.044	.259	.679	.973	1.000
1.708	s.p.	.086	.360	.714	.976	1.000
	n.p.	.086	.389	.799	.988	1.000
2.914	s.p.	.079	.292	.701	.975	1.000
	n.p.	.079	.369	.784	.987	1.000
5.399	s.p.	.050	.280	.697	.975	1.000
	n.p.	.050	.280	.702	.976	1.000

TABLE 3

The power of the sometimes-pool procedure and the never-pool test of the same size, for $n_1 = 14$, $n_2 = 10$, $n_3 = 12$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

θ_{21}	Test	θ_{22}				
		1	2	4	10	50
0.714	s.p.	.013	.232	.708	.971	1.000
	n.p.	.013	.111	.448	.911	1.000
1.045	s.p.	.039	.294	.757	.980	1.000
	n.p.	.039	.240	.657	.969	1.000
1.593	s.p.	.075	.360	.727	.977	1.000
	n.p.	.075	.358	.774	.986	1.000
2.552	s.p.	.082	.311	.704	.975	1.000
	n.p.	.082	.377	.790	.988	1.000
8.066	s.p.	.050	.280	.699	.975	1.000
	n.p.	.050	.230	.702	.976	1.000

TABLE 4.

Illustrating the nature of the approximation to the integral P_1 ($n_1 = 20$ throughout)

	θ_{21}	$n_2 = 6$			$n_2 = 10$			$n_2 = 16$		
		Exact	Approx.	Diff.	Exact	Approx.	Diff.	Exact	Approx.	Diff.
$n_3 = 2$	1.0	.0375	.0375	.0000	.0375	.0375	.0000	.0375	.0375	.0000
	1.5	.0574	.0530	.0044	.0484	.0441	.0043	.0400	.0363	.0037
	2.0	.0621	.0528	.0093	.0444	.0364	.0080	.0304	.0245	.0059
	3.0	.0524	.0396	.0128	.0274	.0190	.0084	.0128	.0084	.0044
	5.0	.0288	.0187	.0101	.0088	.0050	.0038	.0020	.0010	.0010
$n_3 = 6$	1.0	.0375	.0375	.0000				.0375	.0375	.0000
	1.5	.0692	.0623	.0069				.0512	.0445	.0067
	2.0	.0932	.0757	.0175				.0441	.0329	.0112
	3.0	.0856	.0618	.0238				.0208	.0122	.0086
	5.0	.0463	.0305	.0158				.0034	.0016	.0018

For exact integral P_1 , see Eq. (15).

For approximate integral P_1 , see Section 2.4 and Formula (22).

REFERENCES

- [1] T. A. BANCROFT, "On biases in estimation due to the use of preliminary tests of significance," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 190-204.
- [2] T. A. BANCROFT, "Bias due to the omission of independent variables in ordinary multiple regression analysis," (abstract), *Ann. Math. Stat.*, Vol. 21 (1950), p. 142.
- [3] M. S. BARTLETT AND D. G. KENDALL, "The statistical analysis of variance—heterogeneity and the logarithmic transformation," *J. Roy. Stat. Soc., Suppl.*, Vol. 8 (1946), pp. 128-138.
- [4] R. E. BECHHOFFER, "The effect of preliminary tests of significance on the size and power of certain tests of univariate linear hypotheses," unpublished Ph.D. thesis, Columbia University Library, 1951.

- [5] B. M. BENNETT, "Estimation of means on the basis of preliminary tests of significance," *Ann. Math. Stat.*, Vol. 4 (1952), pp. 31-43.
- [6] G. E. P. BOX, "Some theorems on quadratic forms applied in the study of analysis of variance problems I," *Ann. Math. Stat.*, Vol. 25 (1954), p. 290.
- [7] G. E. P. BOX, "Some theorems on quadratic forms applied in the study of analysis of variance problems II," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 484-498.
- [8] H. BOZIVICH, T. A. BANCROFT, AND H. O. HARTLEY, WADC Technical Report (In Press), Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, 1956.
- [9] R. A. FISHER, *The Design of Experiments*, Oliver and Boyd, London, 1937.
- [10] TOSIO KITAGAWA, "Successive process of statistical inference," *Mem. Faculty of Sci.*, Kyusyu University, Series A, Vol. 6, No. 2, 1950.
- [11] M. MERRINGTON AND C. M. THOMPSON, "Tables of percentage points of the inverted Beta (F) distribution," *Biometrika*, Vol. 33 (1943), pp. 73-88.
- [12] FREDERICK MOSTELLER, "On pooling data," *J. Amer. Stat. Assn.*, Vol. 43 (1948), pp. 231-242.
- [13] P. B. PATNAIK, "The non-central χ^2 and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202-232.
- [14] A. E. PAULL, "On a preliminary test for pooling mean squares in the analysis of variance," unpublished Ph.D. thesis, University of North Carolina, 1948.
- [15] A. E. PAULL, "On a preliminary test for pooling mean squares in the analysis of variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 539-556.
- [16] K. PEARSON, editor, *Tables for statisticians and biometricians*, Part 2, London, Biometric Laboratory, 1936.

ON A CLASS OF STOCHASTIC APPROXIMATION PROCESSES¹

BY D. L. BURKHOLDER²

University of North Carolina

1. Summary. We are concerned with the asymptotic behavior of stochastic approximation processes of the Robbins-Monro type [9], the Kiefer-Wolfowitz type [7], and related types. Our main interest is establishing the asymptotic normality, under appropriate conditions, of processes of certain kinds. However, a number of results on convergence with probability one are also obtained as immediate consequences of a theorem needed in the work on asymptotic normality. Our results contain and extend some of the results in this area reported by Blum [1], Chung [3], Hodges and Lehmann [5], and others. In addition, we establish a number of results for cases not previously investigated. For instance, we show that the Kiefer-Wolfowitz processes are asymptotically normal under quite general conditions. The rapidity of convergence depends on the amount that the function M (of Corollary 3.2) departs from symmetry in the neighborhood of the location of the maximum.³ We give results on convergence with probability one and asymptotic normality of stochastic approximation processes useful in connection with the problem of finding the location of the point of inflection of a function. For all cases in which we establish asymptotic normality we also show how the unknown quantities in the variance of the limiting normal distribution can be estimated. These results make possible the construction of asymptotic confidence intervals free of unknowns. Other results, too detailed to be summarized here, are established which also might be of interest in practical applications.

Our method of procedure is as follows. We define a class of stochastic approximation processes, denoted by A_0 , which contains the class of Robbins-Monro processes, the class of Kiefer-Wolfowitz processes, and some other related classes. We first study the class A_0 using methods similar to those used in [1], [3], and [5]. After various results at this level have been obtained, the results for the special cases follow with an economy of effort.

2. Introduction. Let N denote the set of natural numbers and R the set of real numbers. If F is a distribution function and n is in N , let $F^{[n]}$ denote the distribution function of the arithmetic mean of a mutually independent random

Received July 28, 1955; revised April 5, 1956.

¹ This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

² Now at the University of Illinois.

³ The author understands that the asymptotic normality of the Kiefer-Wolfowitz processes has been established independently by C. Derman for the case in which M has the property (along with certain other properties) that there is a $\delta > 0$ such that $M(\theta - \epsilon) = M(\theta + \epsilon)$ for all positive $\epsilon < \delta$, where θ is the location of the maximum of the function M .

variable set of size n , each element of which has the distribution function F . Let M be a function from R into R . For each x in R let $Y(x)$ be a random variable with distribution function $H(\cdot | x)$, such that $EY(x) = M(x)$. Let $\{a_n\}$ be a positive number sequence, let $\{r_n\}$ be a natural number sequence, and let α be in R . Let x_1 be a random variable, and if n is in N let

$$x_{n+1} = x_n - a_n(y_n - \alpha),$$

where y_n is a random variable with conditional distribution function $H^{(r_n)}(\cdot | x_n)$, given $x_1, \dots, x_n, y_1, \dots, y_{n-1}$. The random variable sequence $\{x_n\}$ will be called a *stochastic approximation process of the type A_1* .

This type of process (with $r_n = r_1$ for all n in N) was first developed and used by Robbins and Monro [9]. They proved that if there is a real number θ such that $(x - \theta)(M(x) - \alpha) > 0$ for all real $x \neq \theta$ and if several other conditions are satisfied, then $\{x_n\}$ converges to θ in the mean, which implies, of course, that $\{x_n\}$ converges to θ in probability. Wolfowitz [12] proved convergence in probability under less restrictive conditions. Later, Blum [1], Kallianpur [6], and Kiefer and Wolfowitz (see footnote 2 of [1]) obtained independently and under somewhat different sets of conditions that $\{x_n\}$ converges to θ with probability one. Schmetterer [10, 11] and Kallianpur [6] have reported results on the order of magnitude of $E(x_n - \theta)^2$ under various conditions. Chung [3] has shown that under certain conditions the moments of $a_n^{-1}(x_n - \theta)$ converge to the moments of a normal distribution. This implies, of course, that $a_n^{-1}(x_n - \theta)$ is asymptotically normal. Hodges and Lehmann [5] have relaxed Chung's set of conditions and have obtained asymptotic normality—at the sacrifice, however, of information about the moments of $a_n^{-1}(x_n - \theta)$.

Let M , $Y(\cdot)$ and $H(\cdot | \cdot)$ be as before, let each of $\{a'_n\}$ and $\{c_n\}$ be a positive number sequence, and let $\{r_n\}$ be a natural number sequence. Let x_1 be a random variable, and if n is in N let

$$x_{n+1} = x_n - (a'_n/c_n)(y_{2n-1} - y_{2n}),$$

where y_{2n-1} and y_{2n} are random variables which are conditionally independently distributed according to $H^{(r_n)}(\cdot | x_n - c_n)$ and $H^{(r_n)}(\cdot | x_n + c_n)$, respectively, given $x_1, \dots, x_n, y_1, \dots, y_{2n-2}$. The random variable sequence $\{x_n\}$ will be called a *stochastic approximation process of the type A_2* .

This type of process (with $r_n = r_1$ for all n in N) was first developed and used by Kiefer and Wolfowitz [7]. They proved that if there is a real number θ such that M is increasing for $x < \theta$ and decreasing for $x > \theta$ and if several other conditions are satisfied, then $\{x_n\}$ converges to θ in probability. Later, Blum [1] proved the stronger result that $\{x_n\}$ converges to θ with probability one under less restrictive conditions.

Let M , $Y(\cdot)$, $H(\cdot | \cdot)$, $\{a'_n\}$, $\{c_n\}$, and $\{r_n\}$ be as before. Let x_1 be a random variable, and if n is in N let

$$x_{n+1} = x_n - (a'_n/c_n^2)[y_{2n-1} - (y_{2n-2} + y_{2n})/2],$$

where y_{3n-2} , y_{3n-1} , and y_{3n} are random variables which are conditionally mutually independently distributed according to $H^{(r_n)}(\cdot | x_n - c_n)$, $H^{(r_n)}(\cdot | x_n)$, and $H^{(r_n)}(\cdot | x_n + c_n)$, respectively, given $x_1, \dots, x_n, y_1, \dots, y_{3n-3}$. The random variable sequence $\{x_n\}$ will be called a *stochastic approximation process of the type A_3* . This type of process is useful in connection with the problem of finding the location of the point of inflection of a function, as will be seen later.

It is convenient to study first a type of process slightly more general than any of those defined above. Hence we make the following definition: For each n in N , let R_n be a function from R into R . For each n, x in $N \times R$, let $Z_n(x)$ be a random variable with distribution function $G_n(\cdot | x)$ such that $EZ_n(x) = R_n(x)$. Let $\{a_n\}$ be a positive number sequence. Let x_1 be a random variable, and if n is in N let

$$x_{n+1} = x_n - a_n z_n,$$

where z_n is a random variable with conditional distribution function $G_n(\cdot | x_n)$, given $x_1, \dots, x_n, z_1, \dots, z_{n-1}$. The random variable sequence $\{x_n\}$ will be called a *stochastic approximation process of the type A_0* .

Without confusion, let A_i denote the class of stochastic approximation processes of the type A_i , $i = 0, 1, 2, 3$. It is easy to see that each of A_1, A_2 , and A_3 is a subclass of A_0 .

Throughout this paper, unless otherwise indicated, a passage to the limit will be for $n \rightarrow \infty$. Also, if n is in N , then V_n will denote the function defined by $V_n(x) = \text{var } Z_n(x)$ for x in R , and V will denote the function defined by $V(x) = \text{var } Y(x)$ for x in R .

3. Some results on convergence with probability one. The following theorem is needed in connection with our study of the asymptotic normality of stochastic approximation processes of the type A_0 . However, it also has several immediate applications which we will discuss briefly in this section.

THEOREM 1. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_0 and θ is a real number such that

- (i) there is a function Q from the positive real numbers into N such that if $\epsilon > 0$, $|x - \theta| > \epsilon$, and $n > Q(\epsilon)$, then $(x - \theta)R_n(x) > 0$;
- (ii) $\sup_{n,x} [|R_n(x)| / (1 + |x|)] < \infty$;
- (iii) $\sup_{n,x} V_n(x) < \infty$;
- (iv) if $0 < \delta_1 < \delta_2 < \infty$, then $\sum a_n [\inf_{\delta_1 \leq |x - \theta| \leq \delta_2} |R_n(x)|] = \infty$;
- (v) $\sum a_n^2 < \infty$;
- (vi) if n is in N , then R_n and V_n are Borel measurable.

Then $P \{ \lim x_n = \theta \} = 1$.

We note that condition (vi) implies that if X is a random variable, then each of $R_n(X)$ and $V_n(X)$ is a random variable. Moreover, $V_n(X)$ is integrable, using (iii).

The proof of Theorem 1 is omitted, since the methods used are similar to those used by Blum [1] in the proof of his Theorem 1.

COROLLARY 1.1. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_2 and θ is a real number such that

- (i) M is increasing for $x < \theta$ and decreasing for $x > \theta$;
- (ii) $\sup_x [|M(x-h) - M(x+h)|/(1+|x|)] < \infty$ for some number $h > 0$;
- (iii) V is a bounded, Borel measurable function;
- (iv) if $0 < \delta_1 < \delta_2 < \infty$, then $\inf (|M(x-\epsilon) - M(x+\epsilon)|/\epsilon) > 0$ for $\delta_1 \leq |x - \theta| \leq \delta_2$, $0 < \epsilon < \delta_1$;
- (v) $c_n \rightarrow 0$, $\sum a'_n = \infty$, $\sum (a'_n/c_n)^2 < \infty$.

Then $P\{\lim x_n = \theta\} = 1$.

PROOF. Let $G_n(\cdot | x) = F_n^{(r_n)}(\cdot | x)$, where $F_n(\cdot | x)$ is the distribution function of the sum of two independent random variables, one having the same distribution function as $Y(x - c_n)$, the other having the same distribution function as $-Y(x + c_n)$. Let $Z_n(x)$ be a random variable with distribution function $G_n(\cdot | x)$. Let $R_n(x) = M(x - c_n) - M(x + c_n)$, $z_n = y_{2n-1} - y_{2n}$, and $a_n = a'_n/c_n$. Thus, clearly, $\{x_n\}$ is in A_0 .

Conditions (i) and (ii) of the corollary imply that for each positive number k $\sup [|M(x-\epsilon) - M(x+\epsilon)|/(1+|x|)] < \infty$ for $0 < \epsilon < k$ and x in R . This, in turn, implies that condition (ii) of Theorem 1 is satisfied here.

The reader can easily verify for himself that the other conditions of Theorem 1 are satisfied here, and hence that the desired result holds.

Corollary 1.1 is relevant in connection with the problem of obtaining information about the location θ of the maximum of a regression function M . The above set of conditions is less restrictive than Blum's set of conditions for this case (see Theorem 2 of [1]) in several respects. For instance, regression functions which are parabolic, e.g., $M(x) = -x^2$, and regression functions with bounded value sets, e.g., $M(x) = e^{-x^2}$, are allowed here but are excluded by Blum's set of conditions.

Sometimes information about $M(\theta)$ is desired. The following corollary shows how it can be obtained by using the same data used in the approximation of θ .

COROLLARY 1.1.1. Suppose the conditions of corollary 1.1 hold. If, in addition, M is continuous at θ , then $P\{\lim \sum_{j=1}^n (y_j/n) = M(\theta)\} = 1$.

PROOF. The proof consists of an application of the following special case of Loève's Theorem A [8].

LEMMA 1. If $\{v_n\}$ is a random variable sequence such that $\sum (Ev_n^2/n^2) < \infty$, then $\sum_{j=1}^n [v_j - E(v_j | v_1, \dots, v_{j-1})]/n \rightarrow 0$ with probability one.

Here let $v_{2n-1} = y_{2n-1} - M(x_n - c_n)$ and $v_{2n} = y_{2n} - M(x_n + c_n)$. Then, $Ev_{2n-1}^2 = EV(x_n - c_n)/r_n$ and $Ev_{2n}^2 = EV(x_n + c_n)/r_n$. By condition (iii) of Corollary 1.1, it follows that $\sum (Ev_n^2/n^2) < \infty$. It is easily proved that $E(v_{2n-1} | v_1, \dots, v_{2n-2}) = E(v_{2n} | v_1, \dots, v_{2n-1}) = 0$. Thus, $\sum_{j=1}^n v_j/n \rightarrow 0$ with probability one. By Corollary 1.1, $P\{\lim (x_n \pm c_n) = \theta\} = 1$. Thus, since M is continuous at θ , $P\{\lim M(x_n \pm c_n) = M(\theta)\} = 1$. These facts imply the desired result.

We now consider the implications of Theorem 1 for stochastic approximation processes of the type A_3 . Under certain conditions such processes may be used

to obtain information about the points of inflection of M . However, the results given here will not be stated for the general situation but will be given in terms of the special case of the approximation of the mode of a probability density function. The results for the more general situation follow from Theorem 1 in much the same way.

Throughout the paper, if $0 \leq p \leq 1$, let $B(\cdot | p)$ be the distribution function such that $B(y | p) = 0$ if $y < 0$; $1 - p$ if $0 \leq y < 1$; and 1 if $y \geq 1$.

COROLLARY 1.2. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_2 such that M is a distribution function with associated density function f , $H(\cdot | x) = B(\cdot | M(x))$ for all real x , θ is a real number, and

- (i) f is increasing for $x < \theta$ and decreasing for $x > \theta$;
- (ii) if $0 < \delta_1 < \delta_2 < \infty$, then $\inf(|f(x - \epsilon) - f(x + \epsilon)|/\epsilon) > 0$ for $\delta_1 \leq |x - \theta| \leq \delta_2$, $0 < \epsilon < \delta_1$;
- (iii) $c_n \rightarrow 0$, $\sum a'_n = \infty$, $\sum (a'_n/c_n^2) < \infty$.

Then $P\{\lim x_n = \theta\} = 1$.

PROOF. Let $G_n(\cdot | x) = F_n^{(r_n)}(\cdot | x)$, where $F_n(\cdot | x)$ is the distribution function of the sum of three mutually independent random variables, the first having the same distribution function as $Y(x)$, the second having the same distribution function as $-Y(x - c_n)/2$, and the third having the same distribution function as $-Y(x_n + c_n)/2$. Let $Z_n(x)$ be a random variable with distribution function $G_n(\cdot | x)$. Let $R_n(x) = M(x) - [M(x - c_n) + M(x + c_n)]/2$, $z_n = y_{3n-1} - (y_{3n-2} + y_{3n})/2$, and $a_n = a'_n/c_n^2$. Thus, $\{x_n\}$ is in A_0 . The remainder of the proof is straightforward.

COROLLARY 1.2.1. Suppose the conditions of Corollary 1.2 hold. In addition, suppose f is continuous on an open interval containing θ and $\sum (1/nc_n)^2 < \infty$. Then $P\{\lim \sum_{j=1}^n [(y_{3j} - y_{3j-2})/2c_j n] = f(\theta)\} = 1$.

The proof is similar to the proof of Corollary 1.1.1.

It is not hard to see that Theorem 1 also implies a result on the stochastic approximation of the root of a regression equation. In particular, Theorem 1 implies Blum's [1] Theorem 1 which deals with this case.

We note in passing that our Theorem 1 can be generalized to the case where θ does not exist uniquely. One implication is that if $\{x_n\}$ is a stochastic approximation process of the type A_1 such that M is a distribution function, $H(\cdot | x) = B(\cdot | M(x))$ for all real x , $0 < \alpha < 1$, $\sum a_n^2 < \infty$ and $\sum a_n = \infty$, then there is a random variable x_0 such that $P\{\lim x_n = x_0, \theta_1 \leq x_0 \leq \theta_2\} = 1$, where $\theta_1 = \sup \{x | M(x) < \alpha\}$ and $\theta_2 = \inf \{x | M(x) > \alpha\}$. Thus, information about the quantiles of a distribution function can be obtained by using stochastic approximation methods, even though some of the quantiles may not exist uniquely. The details are given in [2].

4. Asymptotic normality of stochastic approximation processes of the type A_0 .

LEMMA 2. Suppose $\{b_n\}$ is a nonnegative number sequence and each of $\{c_n\}$ and

$\{d_n\}$ is a real number sequence such that

$$b_{n+1} \leq b_n \left[1 - \frac{c_n + o(1)}{n} \right] + \frac{d_n + o(1)}{n^{p+1}} + \frac{O(1)}{n^{q+1}},$$

$$b_{n+1} \leq b_n \left[1 - \frac{c_n}{n} \right] + \frac{d_n}{n^{p+1}} + \frac{O(1)b_n^r}{n^{q+1}},$$

where $\liminf c_n = c > p > 0$, $\overline{\lim} d_n = d \geq 0$, $0 < r < 1$, and $p(1-r) < q$. Then $\lim n^p b_n \leq d/(c-p)$.

This lemma is related to Chung's Lemma 1 [3]. The proof uses Chung's lemma and an induction argument.

LEMMA 3. Let $\{x_n\}$ be a random variable sequence, and if n is in N let each of f_n and g_n be a Borel measurable function from R into R . Suppose $\{f_n\}$ is continuously convergent at the number θ to the number β (that is, if $\{y_n\}$ is a real number sequence with limit θ , then $f_n(y_n) \rightarrow \beta$) and is uniformly bounded. Suppose that if n is in N , then $Eg_n(x_n)$ exists (finitely), and if $\delta > 0$, then

$$E\{|g_n(x_n)| \mid |x_n - \theta| \geq \delta\} P\{|x_n - \theta| \geq \delta\} = o(1)E|g_n(x_n)|.$$

Then $E\{f_n(x_n)g_n(x_n)\} = \beta Eg_n(x_n) + o(1)E|g_n(x_n)|$.

PROOF. Let $K = \sup_{n,x} |f_n(x)|$. By assumption, K is in R . Also, the assumptions imply that $E\{f_n(x_n)g_n(x_n)\}$ exists for all n . Let $\epsilon > 0$. Since $\{f_n\}$ is continuously convergent at θ to β , there is a $\delta > 0$ and an m in N such that if $|x - \theta| < \delta$ and $n > m$, then $|f_n(x) - \beta| < \epsilon$. Thus, for $n > m$

$$\begin{aligned} |E\{(f_n(x_n) - \beta)g_n(x_n)\}| &\leq E\{|f_n(x) - \beta| |g_n(x_n)|\} \\ &\leq \epsilon E\{|g_n(x_n)| \mid |x_n - \theta| < \delta\} P\{|x_n - \theta| < \delta\} \\ &\quad + (K + |\beta|) E\{|g_n(x_n)| \mid |x_n - \theta| \geq \delta\} P\{|x_n - \theta| \geq \delta\} \\ &\leq [\epsilon + o(1)] E|g_n(x_n)|. \end{aligned}$$

The desired result is implied.

We now give several results on the asymptotic normality of stochastic approximation processes of the type A_0 . Although Theorem 2 has implications for the special cases, it is here used only as a step toward Theorem 3. The two theorems will be compared in more detail later.

The methods used in proving Theorem 2 are similar in many respects to the methods used by Chung [3] in his study of the class of Robbins-Monro processes. The main modifications are due chiefly to the fact that usually where Chung is working with a single function, it is necessary here to work with a function sequence. The assumption that there is a real number sequence $\{\mu_n\}$ converging to θ such that $x - \mu_n$ and $R_n(x)$ have the same sign for $x \neq \mu_n$ removes many of the difficulties involved in this approach.

THEOREM 2. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_0 , $\{\mu_n\}$ is a real number sequence, $\{c_n\}$ is a positive number sequence, θ is a real number,

and each of $\beta, T, \bar{T}, \underline{V}, \bar{V}, \sigma^2, \gamma, \xi, c$, and d is a positive number such that

- (i) $R_n(\mu_n) = 0$ for all n ;
- (ii) the function sequence $\{T_n\}$, where for each n in N , $T_n(x) = R_n(x)/c_n(x - \mu_n)$ if $x \neq \mu_n$, $= \beta$ if $x = \mu_n$, is continuously convergent at θ to β and satisfies $\bar{T} \leq T_n(x) \leq \bar{T}$ for all n, x ;
- (iii) $\{V_n\}$ is continuously convergent at θ to σ^2 and satisfies $\underline{V} \leq V_n(x) \leq \bar{V}$ for all n, x ;
- (iv) if r is in N , then $\sup_{n,x} E |Z_n(x) - R_n(x)|^r < \infty$;
- (v) $G_n(y | \cdot)$ is Borel measurable for all n, y ;
- (vi) $\mu_n - \theta = O(n^{-\gamma})$, $0 < \xi \leq \frac{1}{2}$, $\xi < \gamma$, $n^{k+1}a_n \rightarrow c$, $na_n c_n \rightarrow d > \xi/T$;
- (vii) all moments of x_1 exist.

Then, if r is in N ,

$$\lim n^r E(x_n - \theta)^r = \begin{cases} \left[\frac{\sigma^2 c^2}{2\beta d - 2\xi} \right]^{r/2} (r-1)(r-3) \cdots 3 \cdot 1, & \text{if } r \text{ is even,} \\ 0, & \text{if } r \text{ is odd,} \end{cases}$$

which implies that $n^{1/2}(x_n - \theta)$ is asymptotically normal $(0, \sigma^2 c^2 / (2\beta d - 2\xi))$.

PROOF. If n is in N , let $b_n^{(r)} = E(x_n - \theta)^r$, if r is a nonnegative integer, $\beta_n^{(r)} = E |x_n - \theta|^r$, if $r \geq 0$, and $b_n = b_n^{(2)}$. Without going into detail we remark that these expectations and the other expectations written below exist. The necessary integrability of certain functions is assured by boundedness conditions such as (iv), together with the measurability condition (v). That each of $R_n(x)$ and $E |Z_n(x) - R_n(x)|^r$ is Borel measurable in x , for instance, can be seen by considering Stieltjes approximating sums to the integrals involved, $R_n(x) = \int_{-\infty}^{\infty} y dG_n(y | x)$, and so forth.

If r is in N , then

$$(1) \quad b_{n+1}^{(r)} = b_n^{(r)} + \sum_{k=1}^r \binom{r}{k} (-a_n)^k H_k(r, n),$$

where $H_k(r, n) = E[(x_n - \theta)^{r-k} z_n^k]$. By conditions (i) and (ii), $R_n(x) = c_n T_n(x)$ ($x - \mu_n$) for all n, x . If r is in N , then

$$\begin{aligned} (2) \quad H_1(r, n) &= E\{E[(x_n - \theta)^{r-1} z_n | x_n]\} \\ &= E[(x_n - \theta)^{r-1} R_n(x_n)] \\ &= c_n E[T_n(x_n)(x_n - \theta)^r] + c_n(\theta - \mu_n) E[T_n(x_n)(x_n - \theta)^{r-1}]. \end{aligned}$$

Thus, if r is even, then

$$\begin{aligned} (3) \quad H_1(r, n) &\geq c_n \bar{T} b_n^{(r)} - c_n |\theta - \mu_n| \bar{T} \beta_n^{(r-1)} \\ (4) \quad &\geq c_n \bar{T} b_n^{(r)} - c_n |\theta - \mu_n| \bar{T} [b_n^{(r)}]^{(r-1)/r}. \end{aligned}$$

Also, since $2\beta_n^{(r-1)} \leq \beta_n^{(r-2)} + \beta_n^{(r)}$ for $r \geq 2$, we have by (3) and the relation $|\theta - \mu_n| = o(1)$, which is implied by (vi), that if r is even, then

$$(5) \quad H_1(r, n) \geq [c_n \bar{T} + o(c_n)] b_n^{(r)} - c_n |\theta - \mu_n| \bar{T} b_n^{(r-2)} / 2.$$

If r is in N and $r \geq 2$, then

$$(6) \quad H_2(r, n) = E[V_n(x_n)(x_n - \theta)^{r-2}] + c_n^2 E[T_n^2(x_n)(x_n - \theta)^{r-2}(x_n - \mu_n)^2].$$

Since $(x_n - \mu_n)^2 \leq 2(x_n - \theta)^2 + 2(\theta - \mu_n)^2$, we obtain

$$(7) \quad |H_2(r, n)| \leq [\bar{V} + o(1)]\beta_n^{(r-2)} + 2c_n^2 \bar{T}^2 \beta_n^{(r)}.$$

Also, it is easily shown that if each of r and k is in N and $r \geq k$, then

$$(8) \quad |H_k(r, n)| = O(1)\beta_n^{(r-k)} + O(1)c_n^k \beta_n^{(r)}.$$

We shall now prove that (I) if r is a positive number, then there is a positive number B_r such that

$$(9) \quad \lim n^{r/2} \beta_n^{(r)} \leq B_r.$$

Since $[\beta_n^{(r)}]^{1/r}$ is nondecreasing in r for $r > 0$, it suffices to show that (9) holds for each even r .

Consider the case $r = 2$. Using (5) and (7) in (1) gives

$$b_{n+1} \leq b_n \left[1 - \frac{2na_n c_n T + o(1)}{n} \right] + \frac{n^{2\xi+1} a_n^2 \bar{V} + o(1)}{n^{2\xi+1}} + \frac{O(1)}{n^{r+1}}.$$

Using (4) and (7) in (1) gives

$$b_{n+1} \leq b_n \left[1 - \frac{2na_n c_n T + o(1)}{n} \right] + \frac{n^{2\xi+1} a_n^2 \bar{V} + o(1)}{n^{2\xi+1}} + \frac{O(1)b_n^k}{n^{r+1}}.$$

The assumptions of the theorem and the above relations imply, by Lemma 2, that $\lim n^{2\xi} b_n \leq c^2 \bar{V} / 2(Td - \xi)$. Thus, (9) holds for $r = 2$.

Suppose that r is even, $r > 2$, and that (9) holds for each even natural number less than r . Then, of course, $\beta_n^{(k)} = O(n^{-k\xi})$ for each positive number $k \leq r - 2$. Using this fact and (8) gives

$$(10) \quad \sum_{k=3}^r \binom{r}{k} (-a_n)^k H_k(r, n) = o(n^{-r\xi-1}) + o(n^{-1}) b_n^{(r)}.$$

Using the induction hypothesis and substituting (5), (7), and (10) in (1) and (4), (7), and (10) in (1) gives two systems of inequalities which, upon applying Lemma 2, yield the relation $\lim n^{r\xi} b_n^{(r)} \leq (r-1)c^2 \bar{V} B_{r-2} / 2(Td - \xi)$. By induction, (9) holds for each even r . Thus, (I) is proved.

Next, it will be shown that (II) if $\delta > 0$ and r is in N , then

$$E\{|x_n - \theta|^r \mid |x_n - \theta| \geq \delta\} P\{|x_n - \theta| \geq \delta\} = o(1)\beta_n^{(r)}.$$

If $r \geq 0$, $\delta > 0$, $q > 0$, then, by (I),

$$(11) \quad E\{|x_n - \theta|^r \mid |x_n - \theta| \geq \delta\} P\{|x_n - \theta| \geq \delta\} \leq \delta^{-2q/\xi} \beta_n^{(r+2q/\xi)} = o(n^{-q}).$$

Since $x_{n+1} - E(x_{n+1} \mid x_n) = -a_n[z_n - R_n(x_n)]$, we have that $E([x_{n+1} - \theta]^3 \mid x_n) \geq E([x_{n+1} - E(x_{n+1} \mid x_n)]^3 \mid x_n) = a_n^3 E([z_n - R_n(x_n)]^3 \mid x_n) = a_n^3 V_n(x_n) \geq a_n^2 \bar{V}$.

Thus, $(n+1)^{2\lambda} b_{n+1} \geq n^{2\lambda} a_n^2 V = c^2 V + o(1)$, where $\lambda = \xi + \frac{1}{2}$. Therefore, if $r \geq 2$, then

$$(12) \quad \liminf n^{r\lambda} \beta_n^{(r)} \geq \liminf (n^{2\lambda} b_n)^{r/2} \geq (c^2 V)^{r/2}.$$

Furthermore, $n^{2\lambda} \beta_n^{(1)} \geq n^{2\lambda} E\{|x_n - \theta|^2 | |x_n - \theta| < 1\} P\{|x_n - \theta| < 1\} = n^{2\lambda} b_n + o(1)$, using (11). Thus,

$$(13) \quad \liminf n^{2\lambda} \beta_n^{(1)} \geq c^2 V.$$

Relations (11), (12), and (13) imply the assertion (II).

Using (ii), (vi), (I), and (II) in (2) gives, by Lemma 3, that if r is in N , then

$$(14) \quad H_1(r, n) = c_n \beta b_n^{(r)} + c_n o(n^{-r\xi}).$$

Similarly, using (iii), (vi), (I), and (II) in (6) gives that if r is in N and $r \geq 2$, then

$$(15) \quad H_2(r, n) = \sigma^2 b_n^{(r-2)} + o(n^{-(r-2)\xi}).$$

With $r = 1$, using (14) in (1) gives $b_{n+1}^{(1)} = b_n^{(1)} (1 - na_n c_n \beta / n) + o(n^{-\xi-1})$, implying that there is a natural number n_1 such that if $n > n_1$ then $|b_{n+1}^{(1)}| \leq |b_n^{(1)}| (1 - na_n c_n \beta / n) + o(n^{-\xi-1})$. Chung's Lemma 1 implies here that $\lim n^\xi |b_n^{(1)}| = 0$. Thus, $\lim n^\xi b_n^{(1)} = 0$.

With $r = 2$, using (14) and (15) in (1) gives

$$b_{n+1} = b_n \left[1 - \frac{2na_n c_n \beta}{n} \right] + \frac{n^{2\xi+1} a_n^2 \sigma^2 + o(1)}{n^{2\xi+1}}.$$

Applications of Chung's Lemmas 1 and 2 here yield the fact that $\lim n^{2\xi} b_n = \sigma^2 c^2 / 2(\beta d - \xi)$.

A simple inductive argument shows that, in general, if r is in N , then $\lim n^{r\xi} b_n^{(r)}$ exists and is equal to the quantity specified in the conclusion of the theorem. Thus, the theorem is proved.

THEOREM 3. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_0 , $\{\mu_n\}$ is a real number sequence, $\{c_n\}$ is a positive number sequence, θ is a real number, and each of n_0 , β , σ^2 , ϵ_0 , γ , ξ , c , and d is a positive number such that conditions (ii) and (iv) of Theorem 1 and (v) of Theorem 2 are satisfied and

- (i) if $n > n_0$ and $x \neq \mu_n$, then $(x - \mu_n) R_n(x) > 0$ and $R_n(\mu_n) = 0$;
- (ii) $\{T_n\}$, defined as in Theorem 2, is continuously convergent at θ to β ;
- (iii) $\{V_n\}$ is uniformly bounded and is continuously convergent at θ to σ^2 ;
- (iv) if r is in N , then $\sup E|Z_n(x) - R_n(x)|^r < \infty$ for $|x - \theta| < \epsilon_0$, $n > n_0$,
- (v) $\mu_n - \theta = O(n^{-\gamma})$, $0 < \xi \leq \frac{1}{2}$, $\xi < \gamma$, $n^{\xi+1} a_n \rightarrow c$, $na_n c_n \rightarrow d > \xi/\beta$.

Then $n^\xi(x_n - \theta)$ is asymptotically normal $(0, \sigma^2 c^2 / 2(\beta d - \xi))$.

The conclusion of Theorem 3 is not as strong, of course, as the conclusion of Theorem 2. On the other hand, the set of conditions of Theorem 3 is considerably less restrictive. Conditions (ii) and (iv) of Theorem 1 are much weaker restrictions on $\{R_n\}$ than the assumptions involving T and \bar{T} of Theorem 2. In Theorem

3, assumptions about the moments of $G_n(\cdot | x)$ higher than the second are made only for x in a neighborhood of θ . No assumptions about the moments of the random variable x_1 are made, and so forth.

In the light of the above, the special cases will be considered here in terms of Theorem 3 only.

The truncation device used in the proof was introduced by Hodges and Lehmann [5] in their work on the class of Robbins-Monro processes. With it they were able to relax a condition of Chung's Theorem 9 [3] analogous to the one involving \bar{T} in condition (ii) of Theorem 2.

PROOF. Let T, \bar{T}, V , and \bar{V} be positive numbers satisfying $\xi/d < T < \beta < \bar{T}$ and $V < \sigma^2 < \bar{V}$. By assumptions (ii) and (iii), there is a positive number $\delta < \epsilon_0$ and a natural number $n_1 > n_0$ such that if x is in $I = [\theta - \delta, \theta + \delta]$ and $n > n_1$, then $T < T_n(x) < \bar{T}$ and $V < V_n(x) < \bar{V}$. Let y be in R and let $\epsilon > 0$. The conditions of Theorem 1 are satisfied here. Thus, $P\{\lim x_n = \theta\} = 1$, which implies there is a natural number $m > n_1$ such that

$$(16) \quad P\{\sup_n |x_{m+n} - \theta| < \delta\} \geq 1 - \epsilon.$$

For each n, x in $N \times R$, let $Z'_n(x) = Z_n(x)$, if x is in I ; $Z_n(\theta - \delta) - R_n(\theta - \delta) + \beta c_n(x - \mu_n)$ if x is not in I ; let $G'_n(\cdot | x)$ be the distribution function of $Z'_n(x)$ and let $R'_n(x) = EZ'_n(x)$. Let x'_1 be the random variable such that $x'_1 = x_{m+1}$ if x_{m+1} is in I ; 0, otherwise. For each n in N , let $x'_{n+1} = x'_n - a_{m+n}z'_n$, where z'_n is a random variable with conditional distribution function $G'_{m+n}(\cdot | x'_1)$, given $x'_1, \dots, x'_n, z'_1, \dots, z'_{n-1}$ and is such that $P\{z'_n = z_{m+n} | A\} = 1$, where $A = \{x'_1, \dots, x'_n, z'_1, \dots, z'_{n-1} = x_{m+1}, \dots, x_{m+n}, z_{m+1}, \dots, z_{m+n-1}\}$.

From (16) and the definition of $\{x'_n\}$, we have that if n is in N , then $P\{x'_n \neq x_{m+n}\} < \epsilon$, implying that

$$|P\{(m+n)^\xi(x'_n - \theta) \leq y\} - P\{(m+n)^\xi(x_{m+n} - \theta) \leq y\}| < \epsilon.$$

It is clear that $\{x'_n\}$ is a stochastic approximation process of the type A_0 and satisfies the conditions of Theorem 2. Therefore, $n^\xi(x'_n - \theta)$ and, consequently, $(m+n)^\xi(x'_n - \theta)$, is asymptotically normal $(0, \sigma^2 c^2 / 2(\beta d - \xi))$. Denoting this normal distribution by D , we have that there is an n_2 in N such that if $n > n_2$ then $|P\{(m+n)^\xi(x'_n - \theta) \leq y\} - D(y)| < \epsilon$ and, hence, $|P\{(m+n)^\xi(x_{m+n} - \theta) \leq y\} - D(y)| < 2\epsilon$. The desired result is implied.

THEOREM 4. Suppose the conditions of Theorem 3 hold. Then $n^{1/2}R_n(x_n)$ is asymptotically normal $(0, \sigma^2 \beta^2 d^2 / 2(\beta d - \xi))$.

PROOF. For all $n > n_0$, $n^{1/2}R_n(x_n) = n^\xi(x_n - \theta)T_n(x_n)n^{1-\xi}c_n + o(1)T_n(x_n)$. By Cramér's theorem ([4], p. 254), the desired conclusion follows from the conditions and conclusion of Theorem 3.

5. Asymptotic confidence intervals free of unknowns. We note that the quantities σ^2 and β , each of which would probably be unknown in most practical situations, appear in the variance of each of the asymptotic distributions given above. In order to be able to construct sequences of estimates of these quantities

of the sort that would make it possible to obtain asymptotic confidence intervals for θ free of unknowns, we proceed as follows:

With each stochastic approximation process of the type A_0 we associate a random variable sequence $\{z_n\}$ and a positive number sequence $\{\pi_n\}$ such that for each n in N , z_n is a random variable, conditionally independent of z_n , with conditional distribution function $G_n(\cdot | x_n + \pi_n)$, given $x_1, \dots, x_n, z_1, \dots, z_{n-1}$.

Also, to each $\{x_n\}$ in A_i we associate a random variable sequence $\{\tilde{y}_n\}$ and a positive number sequence $\{\pi_n\}$ in an analogous manner, $i = 1, 2, 3$. For instance, if $\{x_n\}$ is in A_2 , then for each n in N , \tilde{y}_{2n-1} and \tilde{y}_{2n} are random variables which are conditionally independently distributed according to $H^{(r_n)}(\cdot | x_n - c_n + \pi_n)$ and $H^{(r_n)}(\cdot | x_n + c_n + \pi_n)$, respectively, and which are conditionally independent of y_{2n-1} and y_{2n} , given $x_1, \dots, x_n, y_1, \dots, y_{2n-2}$.

THEOREM 5. Suppose the conditions of Theorem 3 hold. Then $s_n^2 = \sum_{i=1}^n z_i^2/n \rightarrow \sigma^2$ with probability one. If, in addition, $\pi_n \rightarrow 0$ and $1/\pi_n = O(n^{\xi_0})$ where $0 < \xi_0 < \xi$, then $t_n = \sum_{i=1}^n [(z_i - z_j)/c_j \pi_j n] \rightarrow \beta$ with probability one, which implies that $n^{\xi} |2t_n d - 2\xi|^{1/2} (x_n - \theta)/c s_n$ is asymptotically normal $(0, 1)$.

Thus, asymptotic confidence intervals for θ free of unknowns can be constructed.

PROOF. The last remark follows, by Cramér's theorem ([4], p. 254), from the conclusion of Theorem 3 and the fact that $c^2 s_n^2 / (2(t_n d - \xi)) \rightarrow c^2 \sigma^2 / 2 (\beta d - \xi)$ with probability one. (Convergence in probability is, of course, enough.)

We will first show that (I) if, in addition to the above conditions, the conditions of Theorem 2 hold, then $s_n^2 \rightarrow \sigma^2$ with probability one and $(x_n - \theta)/\pi_n \rightarrow 0$ with probability one.

Let $v_n = z_n^2 - V_n(x_n) - R_n^2(x_n)$. Then $v_n^2 \leq 3 [z_n^4 + V_n^2(x_n) + R_n^4(x_n)]$ and $E v_n^2 = O(1)$, where the latter fact is obtained by using the results of calculations made in the proof of Theorem 2. Hence, $\sum (E v_n^2 / n^2) < \infty$. Also, $E(v_n | v_1, \dots, v_{n-1}) = 0$ for all n . Thus, by Lemma 1, $\sum_{i=1}^n v_i / n \rightarrow 0$ with probability one. Since $P\{\lim x_n = \theta\} = 1$, $V_n(x_n) + R_n^2(x_n) \rightarrow \sigma^2$ with probability one, implying that $P\{\lim s_n^2 = \sigma^2\} = 1$.

Let $r > 1/(\xi - \xi_0)$. Then by Techebycheff's inequality and relation (9) we have that $P\{|x_n - \theta|/\pi_n \geq \epsilon\} \leq \beta_n^{(r)}/\pi_n^r \epsilon^r = O(n^{-r(\xi - \xi_0)})$ for each $\epsilon > 0$. By the Borel-Cantelli lemma, the latter part of (I) follows.

Using (I) and employing the truncation device used in the proof of Theorem 3, it is easily shown that (II) under the conditions of Theorem 5, $s_n^2 \rightarrow \sigma^2$ with probability one and $(x_n - \theta)/\pi_n \rightarrow 0$ with probability one.

An application of Lemma 1 gives that $t_n = \sum_{i=1}^n [(R_j(x_j + \pi_j) - R_j(x_j))/c_j \pi_j n] \rightarrow 0$ with probability one. For $n > n_0$, $[R_n(x_n + \pi_n) - R_n(x_n)]/c_n \pi_n = [T_n(x_n + \pi_n) - T_n(x_n)](x_n - \theta + \theta - \mu_n)/\pi_n + T_n(x_n + \pi_n)$. The first term of the right-hand side converges to 0 with probability one and the second term converges to β with probability one, using (II), the fact that $(\theta - \mu_n)/\pi_n = O(n^{-\gamma + \xi_0})$, where $\xi_0 < \xi < \gamma$, and condition (ii) of Theorem 3. This completes the proof.

6. Approximation of a root of a regression equation. We will now sketch some of the implications of the above results for processes of the type A_1 , A_2 , and A_3 .

COROLLARY 3.1. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_1 , θ is a real number and each of α_1 , σ^2 , ϵ_0 , c , and r_0 is a positive number such that

- (i) if $x \neq \theta$, then $(x - \theta)(M(x) - \alpha) > 0$;
- (ii) M is differentiable at θ and $M'(\theta) = \alpha_1$;
- (iii) $\sup_x [|M(x)|/(1 + |x|)] < \infty$;
- (iv) if $0 < \delta_1 < \delta_2 < \infty$ then $\inf |M(x) - \alpha| > 0$ for $\delta_1 \leq |x - \theta| \leq \delta_2$;
- (v) V is bounded and is continuous at θ , and $V(\theta) = \sigma^2$;
- (vi) if r is in N , then $\sup E|Y(x) - M(x)|^r < \infty$ for $|x - \theta| < \epsilon_0$;
- (vii) $H(y|\cdot)$ is Borel measurable for each y in R ;
- (viii) $na_n \rightarrow c > 1/2\alpha_1$, $r_n \rightarrow r_0$.

Then $n^{1/2}(x_n - \theta)$ is asymptotically normal $(0, \sigma^2 c^2 / r_0(2\alpha_1 c - 1))$.

The first results of the above kind were obtained by Chung [3]. Later, Hodges and Lehmann [5] modified a result of Chung's and obtained a weaker conclusion, asymptotic normality without knowledge of moments, under weaker conditions. Corollary 3.1 is essentially the same as the result of Hodges and Lehmann. Minor differences are as follows: here the possibility that $\liminf_{|x| \rightarrow \infty} |M(x) - \alpha| = 0$ is allowed, and assumptions about the moments of $H(\cdot|x)$ higher than the second are made only for x in a neighborhood of θ . The main reason for stating Corollary 3.1 here is for later use.

The proof of Corollary 3.1 consists of showing that $\{x_n\}$ is in A_0 and that the conditions of Theorem 3 are satisfied. Except for establishing that condition (vii) of the corollary implies that condition (v) of Theorem 2 is satisfied, it is trivial. The exceptional part is taken care of at once by the following lemma.

LEMMA 4. For each x in R let $Y_1(x), \dots, Y_m(x)$ be mutually independent random variables where $H_j(\cdot|x)$ is the distribution function of $Y_j(x)$, $j = 1, \dots, m$, and $G(\cdot|x)$ is the distribution function of $\sum_{j=1}^m Y_j(x)$. Suppose that if y is in R , then $H_j(y|\cdot)$ is Borel measurable, $j = 1, \dots, m$. Then for each y in R , $G(y|\cdot)$ is Borel measurable.

The method of proof is as follows. For each n in N express

$$\tilde{G}(y + 1/n|x) - \tilde{G}(y - n|x),$$

where $2\tilde{G}(v|x) = G(v - |x) + G(v + |x)$ for v in R , in terms of Lévy's inversion formula for characteristic functions. Using Stieltjes approximating sums and the assumptions of the lemma, it is not hard to show that

$$\tilde{G}(y + 1/n|\cdot) - \tilde{G}(y - n|\cdot)$$

is Borel measurable. Thus, since

$$\tilde{G}(y + 1/n|\cdot) - \tilde{G}(y - n|\cdot) \rightarrow G(y|\cdot),$$

$G(y|\cdot)$ is Borel measurable.

COROLLARY 4.1. Under the conditions of Corollary 3.1, $n^{1/2}(M(x_n) - \alpha)$ is asymptotically normal $(0, \sigma^2 \alpha_1^2 c^2 / r_0(2\alpha_1 c - 1))$.

COROLLARY 5.1. Suppose the conditions of Corollary 3.1 hold. Then $s_n^2 = r_0 \sum_{j=1}^n (y_j - \alpha)^2 / n \rightarrow \sigma^2$ with probability one. If, in addition, $\pi_n \rightarrow 0$ and $1/\pi_n = O(n^{\xi_0})$, where $0 < \xi_0 < \frac{1}{2}$, then $t_n = \sum_{j=1}^n [(\bar{y}_j - y_j)/\pi_j n] \rightarrow \alpha_1$ with probability one, which implies that $n^{\frac{1}{2}} r_0^{\frac{1}{2}} |2t_n c - 1|^{\frac{1}{2}} (x_n - \theta)/(cs_n)$ is asymptotically normal $(0, 1)$.

Of course, under all the conditions of the above corollary, $n^{\frac{1}{2}} r_0^{\frac{1}{2}} |2t_n c - 1|^{\frac{1}{2}}$. $(M(x_n) - \alpha)/cs_n |t_n|$ is asymptotically normal $(0, 1)$, also. The results for $M(x_n) - \alpha$ may be of some interest in practical applications.

7. Approximation of the location of the maximum of a regression function.

If θ is a real number let \mathfrak{M}_θ be the set such that M is in \mathfrak{M}_θ if and only if M is a function from R into R such that either (i) M is increasing for $x \leq \theta$ and decreasing for $x > \theta$, or (ii) M is increasing for $x < \theta$ and decreasing for $x \geq \theta$.

If M is in \mathfrak{M}_θ , let $\mu_M = \mu$ be the function from the positive numbers into R such that if $\epsilon > 0$ then $[x - \mu(\epsilon)][M(x - \epsilon) - M(x + \epsilon)] > 0$ for all $x \neq \mu(\epsilon)$.

LEMMA 5. If M is in \mathfrak{M}_θ , then the function μ exists and $|\mu(\epsilon) - \theta| \leq \epsilon$ for each $\epsilon > 0$.

PROOF. Suppose M satisfies (i) of the definition of \mathfrak{M}_θ . (The other case is handled similarly.) Then $M(x - \epsilon) - M(x + \epsilon)$ is negative for $x \leq \theta - \epsilon$, positive for $x > \theta + \epsilon$, and increasing in x for x in $(\theta - \epsilon, \theta + \epsilon)$. This implies the desired result.

If M is in \mathfrak{M}_θ then we will say that M is η -locally-even at θ if and only if $\eta \geq 0$ and $\mu(\epsilon) - \theta = O(\epsilon^{1+\eta})$ as $\epsilon \rightarrow 0$.

By Lemma 5, if M is in \mathfrak{M}_θ , then M is at least 0-locally-even at θ . Lemma 6, below, indicates that each function in a fairly large subset of \mathfrak{M}_θ is 1-locally-even at θ . If $M(\theta - \epsilon) = M(\theta + \epsilon)$ for each positive $\epsilon < \delta$ then $\mu(\epsilon) = \theta$ for $0 < \epsilon < \delta$, implying that M is η -locally-even at θ for each $\eta \geq 0$. How locally-even M is at θ is an important factor affecting the asymptotic behavior of type A_2 processes.

LEMMA 6. If M is in \mathfrak{M}_θ and δ is a positive number such that the first three derivatives of M exist on $I = [\theta - \delta, \theta + \delta]$, $M''(\theta) \neq 0$, and either $M^{(3)}$ is continuous on I or is of bounded variation on I , then M is 1-locally-even at θ .

PROOF. In either case $M(x) = M(\theta) + M''(\theta)(x - \theta)^2/2 + O(|x - \theta|^3)$ as $x \rightarrow \theta$. Therefore, for $0 < \epsilon < \delta/2$, $0 = M(\mu(\epsilon) - \epsilon) - M(\mu(\epsilon) + \epsilon) = -2M''(\theta)\epsilon(\mu(\epsilon) - \theta) + O(\epsilon^3)$ as $\epsilon \rightarrow 0$. The desired result follows.

COROLLARY 3.2. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_2 , θ is a real number, and each of α_2 , σ^2 , ϵ_0 , ξ , c , d , and r_0 is a positive number such that conditions (ii) and (iv) of Corollary 1.1 are satisfied and

- (i) M is in \mathfrak{M}_θ , the first two derivatives of M exist on an open interval containing θ , M'' is continuous at θ , and $M''(\theta) = -\alpha_2$;
- (ii) V is bounded and is continuous at θ , and $V(\theta) = \sigma^2$;
- (iii) if r is in N , then $\sup E |Y(x) - M(x)|^r < \infty$ for $|x - \theta| < \epsilon_0$;
- (iv) $H(y | \cdot)$ is Borel measurable for each y in R ;

(v) M is η -locally-even at θ , and $0 < \xi < \frac{1}{2} - [1/(4 + 2\eta)]$, $n^{\xi+1}a'_n/c_n \rightarrow c$, $na'_n \rightarrow d > \xi/2\alpha_2$, $r_n \rightarrow r_0$.

Then $n^\xi(x_n - \theta)$ is asymptotically normal $(0, \sigma^2 c^2 / r_0 (2\alpha_2 d - \xi))$.

Here the μ_n of Theorem 3 is, of course, $\mu(c_n)$.

COROLLARY 3.2.1. Under the conditions of Corollary 3.2,

$$\lim P\{2n^{2\xi}[M(\theta) - M(x_n)] \leq x\alpha_2 K^2\} = \int_0^\infty (2\pi t)^{-1/2} e^{-t/2} dt$$

for each positive number x , where $K^2 = \sigma^2 c^2 / r_0 (2\alpha_2 d - \xi)$.

PROOF. The assumptions imply that

$$M(x) = M(\theta) - \alpha_2(x - \theta)^2/2 + o(|x - \theta|^2)$$

as $x \rightarrow \theta$. Thus,

$$2n^{2\xi}[M(\theta) - M(x)]/\alpha_2 K^2 = [n^\xi(x - \theta)/K]^2 + o(1)n^{2\xi}(x - \theta)^2.$$

The desired result follows easily from the conclusion of Corollary 3.2.

COROLLARY 3.1.1. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_2 , θ is a real number, and each of ϵ , α_1 , σ^2 , c , and r_0 is a positive number such that condition (iv) of Corollary 3.2 holds, and

(i) M is in \mathfrak{M}_θ , M is differentiable at $\theta_1 - \epsilon$ and $\theta_1 + \epsilon$, and

$$M'(\theta_1 - \epsilon) - M'(\theta_1 + \epsilon) = 2\alpha_1,$$

where θ_1 denotes $\mu(\epsilon)$;

(ii) $\sup_x [|M(x - \epsilon) - M(x + \epsilon)| / (1 + |x|)] < \infty$;

(iii) V is bounded and is continuous at $\theta_1 - \epsilon$ and $\theta_1 + \epsilon$, and

$$V(\theta_1 - \epsilon) + V(\theta_1 + \epsilon) = 2\sigma^2;$$

(iv) there is an open real number set J containing $\theta_1 - \epsilon$ and $\theta_1 + \epsilon$ such that if r is in N , then $\sup E |Y(x) - M(x)|^r < \infty$ for x in J ;

(v) $na'_n \rightarrow c > 1/4\alpha_1$, $r_n \rightarrow r_0$, and if n is in N , then $c_n = \epsilon$.

Then $n^{1/2}(x_n - \mu(\epsilon))$ is asymptotically normal $(0, 2\sigma^2 c^2 / r_0 (4\alpha_1 c - 1))$.

Here the conclusion is given in terms of $\mu(\epsilon)$ rather than in terms of θ . Nevertheless, the inequality $|\mu(\epsilon) - \theta| \leq \epsilon$ holds. Also, if $M(\theta - \epsilon) = M(\theta + \epsilon)$, then $\mu(\epsilon) = \theta$. The main point of the above result is that n^ξ in the conclusion of Corollary 3.2 is replaced by $n^{1/2}$ in the above conclusion. We remember that $\xi < \frac{1}{2}$ in Corollary 3.2. It is desirable, of course, to have the exponent of n as large as possible.

The class of processes satisfying the conditions of 3.1.1 is a subclass of A_1 . Once this is observed, the above result follows easily from Corollary 3.1.

We note that results analogous to Corollary 5.1 can easily be written down for the cases considered in Corollaries 3.2 and 3.1.1.

8. Approximation of the mode of a density function. Again, instead of considering the general point-of-inflection problem, we limit ourselves to the special case of the approximation of the mode of a density function. Results for the more general situation can be obtained in the same way.

If M is a distribution function with associated density function f and f is in \mathfrak{M}_θ , let $\lambda_\epsilon = \lambda$ be the function from the positive numbers into R such that if $\epsilon > 0$, then $[x - \lambda(\epsilon)] [2M(x) - M(x - \epsilon) - M(x + \epsilon)] > 0$ for all $x \neq \lambda(\epsilon)$.

LEMMA 7. If M and f are as above, then the function λ exists and $|\lambda(\epsilon) - \theta| \leq \epsilon$ for each $\epsilon > 0$.

PROOF. Let $M_\epsilon(x) = \int_{x-\epsilon}^{x+\epsilon} f(t) dt$ for all x, ϵ . Then for each $\epsilon > 0$, M_ϵ is in $\mathfrak{M}_{\theta f(\epsilon)}$. Also,

$$2M(x) - M(x - \epsilon) - M(x + \epsilon) = M_{\epsilon/2}(x - \epsilon/2) - M_{\epsilon/2}(x + \epsilon/2).$$

Let $\lambda(\epsilon) = \mu_{M_{\epsilon/2}}(\epsilon/2)$. The desired result follows by using Lemma 5.

LEMMA 8. If M and f are as above and, in addition, there is a $\delta > 0$ such that the first three derivatives of f exist on $I = [\theta - \delta, \theta + \delta]$, $f''(\theta) \neq 0$ and either $f^{(3)}$ is continuous on I or is of bounded variation on I , then $\lambda(\epsilon) - \theta = O(\epsilon^2)$ as $\epsilon \rightarrow 0$.

The proof is similar to that of Lemma 6.

COROLLARY 3.3. Suppose $\{x_n\}$ is a stochastic approximation process of the type A_1 , where M is a distribution function with associated density function f and $H(\cdot | x) = B(\cdot | M(x))$ for x in R , θ is a real number, η is a nonnegative number, and each of α_3 , ξ , c , d , and r_0 is a positive number such that condition (ii) of Corollary 1.2 is satisfied, and

(i) f is in \mathfrak{M}_θ , the first two derivatives of f exist on an open interval I containing θ , f'' is continuous on I , and $f''(\theta) = -\alpha_3$;

(ii) $\lambda(\epsilon) - \theta = O(\epsilon^{1+\eta})$ as $\epsilon \rightarrow 0$, $0 < \xi < \frac{1}{2} - 1/(3 + \eta)$, $n^{\xi+1}d'_n/c_n^2 \rightarrow c$, $na'_n \rightarrow d > 2\xi/\alpha_3$, $r_n \rightarrow r_0$.

Then $n^\xi(x_n - \theta)$ is asymptotically normal $(0, 3\sigma^2 c^2 / 2r_0(\alpha_3 d - 2\xi))$, where $\sigma^2 = M(\theta)[1 - M(\theta)]$.

The asymptotic distribution of $2n^{2\xi}[f(\theta) - f(x_n)]$ can be derived using methods similar to those used in proving Corollary 3.2.1. Also, a corollary to Theorem 5 for this case can be written down.

9. Acknowledgments. The author is pleased to acknowledge his indebtedness to Professor Wassily Hoeffding for helpful advice in connection with this work and to J. L. Hodges, Jr., and E. L. Lehmann for the opportunity of seeing their paper in manuscript form.

REFERENCES

- [1] J. R. BLUM, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382-386.
- [2] D. L. BURKHOLDER, "On a certain class of stochastic approximation processes," unpublished thesis, University of North Carolina, 1955.

- [3] K. L. CHUNG, "On a stochastic approximation method," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 463-483.
- [4] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
- [5] J. L. HODGES, JR., AND E. L. LEHMANN, "Two approximations to the Robbins-Monro process," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, to be published.
- [6] G. KALLIANPUR, "A note on the Robbins-Monro stochastic approximation method," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 386-388.
- [7] J. KIEFER AND J. WOLFOWITZ, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462-466.
- [8] M. LOÈVE, "On almost sure convergence," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 279-303.
- [9] H. ROBBINS AND S. MONRO, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400-407.
- [10] L. SCHMETTERER, "Bemerkungen zum Verfahren der stochastischen Iteration," *Österreich. Ing.-Arch.*, Vol. 7 (1953), pp. 111-117.
- [11] L. SCHMETTERER, "Zum Sequentialverfahren von Robbins und Monro," *Monatsh. Math.*, Vol. 58 (1954), pp. 33-37.
- [12] J. WOLFOWITZ, "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 457-461.

ON SEQUENTIAL DESIGNS FOR MAXIMIZING THE SUM OF n OBSERVATIONS¹

BY R. N. BRADT, S. M. JOHNSON, AND S. KARLIN

1. Introduction. An important simple type of sequential design problem is as follows: We have two binomial random variables, X and Y , having parameters under the two hypotheses, H_1 and H_2 , given by

	X	Y
(ξ) H_1	p	q
($1 - \xi$) H_2	q	p

where ξ is the a priori probability that H_1 is true. We wish to maximize the sum of n observations. The procedure for selecting an X or Y observation at each stage, of course, takes account of all the previous history.

A more realistic version of the design problem deals with the situation such that X and Y have parameters p and q , respectively, where an a priori distribution $F(p, q)$ is known. The problem holds interest for several reasons. It would appear to be one of the simplest problems in the sequential design of an experiment that can be posed; hence its analysis is a step towards obtaining a body of information relative to *specific* sequential design problems. It has not only this general interest but also, as it stands, it has applications in particular problems such as learning theory, biology, and medicine; see [1], for instance, in which applications in the latter two fields may be found. A discussion of problems of this general variety and of certain strategies has been published by Robbins [2]. More immediately, in the final section of this paper it is shown that the solution to the problem in which p has a priori distribution F and q is assumed known, explicitly obtained in Section 4, yields directly the solution of a problem in industrial inspection.

The type of problem known as the "Two-armed Bandit" is a special case of the preceding. In its "classical" formulation (whence the name), we have a slot machine with two arms, an X -arm and a Y -arm. When either arm is pulled, the machine pays off either one unit or nothing; and the probability of winning with one arm is p , and, with the other, q . A priori it is unknown which is which, but the probability ξ that it is the X -arm which has probability p of success is assumed known. One is allowed n plays, and a sequential design, or strategy, is desired which will maximize the expected winnings.

We shall use here for intuitive concreteness the gambling interpretation and terminology.

It has been conjectured for this problem that the optimal strategy is S_1 : on

Received Sept. 28, 1954.

¹ This work was supported in part by an Office of Naval Research contract at Stanford University.

each play choose the arm having, at that time, the maximum expected probability of paying off, i.e., play each time as though there were but one play remaining. This conjecture has been verified to hold for $n \leq 8$.

The "Two-armed Bandit" problem can be generalized in two directions. The random variables may have distributions other than binomial. Sufficient conditions that S_1 be optimal are given and it is shown that, for the binomial case following S_1 , the expected winnings per play tend to $\max[p, q]$ as $n \rightarrow \infty$.

The other direction of generalization was the problem as originally introduced: X and Y are binomial with parameters p and q having a priori distribution $F(p, q)$. In Section 3 it is shown that several properties which one intuitively expects the optimum strategy to possess are not, in general, characteristics of the optimal strategy—e.g., for p and q independently distributed, if only n is sufficiently large, S_1 is not optimal; the optimal strategy may not stay on a winner; and the expected gain on the r -th play is not necessarily a nondecreasing function of r . Also, S_k , the strategy which maximizes the expected winnings over the next k plays, is not always an improvement over S_{k-1} .

In Section 4, the parameter q and the a priori distribution $F(p)$ are assumed known. In this case the optimal strategy is determined *explicitly* and is shown to have those intuitive properties which have been previously noted not to be general characteristics of optimal strategies. These results are applied, in the final section, to obtain the optimal procedure in a certain industrial inspection problem.

2. The "Two-armed Bandit." 2.1. The statistical problem which goes under this general title is that of finding a design which will maximize the sum of n independent observations in the following situation: Let X and Y be real-valued random variables having cdf's F_i and G_i , respectively, under hypothesis H_i ($i = 1, 2$) and ζ be the a priori probability that H_1 is the true hypothesis. The problem is to devise a sequential design which will maximize the expected value of the sum of n observations, each of which is to be an observation either of X or of Y .

Let f_i and g_i be the densities corresponding to F_i and G_i with respect to the measure ψ . Let $W_n(\zeta, S^*)$ denote the expected value of the sum of the n observations if ζ is the a priori probability for H_1 and the design, S^* , is used. If one observed X first and then continued for $n - 1$ steps following the optimal rule S^* , then the expected sum would be

$$\begin{aligned} A_n = & \zeta \int_{-\infty}^{\infty} t f_1(t) d\psi + (1 - \zeta) \int_{-\infty}^{\infty} t f_2(t) d\psi \\ (2.1.1) \quad & + \int_{-\infty}^{\infty} W_{n-1} \left(\frac{\zeta f_1(t)}{\zeta f_1(t) + (1 - \zeta) f_2(t)}, S^* \right) [\zeta f_1(t) + (1 - \zeta) f_2(t)] d\psi. \end{aligned}$$

Similarly, if Y were observed first and the optimal rule followed for the re-

maining $n - 1$ steps, the expected sum would be

$$(2.1.2) \quad B_n = \xi \int_{-\infty}^{\infty} t g_1(t) d\psi + (1 - \xi) \int_{-\infty}^{\infty} t g_2(t) d\psi \\ + \int_{-\infty}^{\infty} W_{n-1} \left(\frac{\xi g_1(t)}{\xi g_1(t) + (1 - \xi) g_2(t)}, S^* \right) [\xi g_1(t) + (1 - \xi) g_2(t)] d\psi.$$

Hence, $W_n(\xi, S^*) = \max(A_n, B_n)$.

A natural design to be considered is that which requires that one maximize step by step—i.e., after the j -th observation, the a posteriori probability, ξ_j , is computed; and at the next step, the random variable corresponding to the maximum of

$$\int_{-\infty}^{\infty} t [\xi_j f_1(t) + (1 - \xi_j) f_2(t)] d\psi$$

and

$$\int_{-\infty}^{\infty} t [\xi_j g_1(t) + (1 - \xi_j) g_2(t)] d\psi$$

is observed. Denote this stepwise maximization design by S_1 .

THEOREM 2.1. *If the likelihood ratios f_2/f_1 and g_2/g_1 have the same distributions under H_1 and also under H_2 , then S_1 is the optimal design.*

PROOF. Since

$$(1) \quad \xi_1 = \begin{cases} \frac{1}{1 + \frac{1 - \xi}{\xi} \frac{f_2(x)}{f_1(x)}} & \text{if } X \text{ is observed first,} \\ \frac{1}{1 + \frac{1 - \xi}{\xi} \frac{g_2(x)}{g_1(x)}} & \text{if } Y \text{ is observed first,} \end{cases}$$

and the likelihood ratios have the same distributions, the distribution of ξ_1 is independent of which random variable is observed first. Hence, the expected value of the optimal yield from the last $(n - 1)$ steps is independent of the choice for the first step. One can, therefore, maximize the expected sum of n observations by choosing at the first step the random variable having the larger expected value and continuing with the optimal design for the remaining steps.

Since all the random variables are assumed to be independent, the same argument shows that, given ξ_j , it is optimal to follow S_1 for the $(j + 1)$ -st step.

An example in which the likelihood ratios are distributed alike is:

	X	Y	
H_1	$N(0, 1)$	$N(\nu, 1)$	
H_2	$N(\mu, 1)$	$N(0, 1)$	$(\mu > 0, \nu > 0),$

with $\mu = \nu$. However, it can be shown that for $n = 2$ and $(1 - \zeta)\mu = \zeta\nu$, S_1 is optimal *only* if $\mu = \nu$.

2.2. A special case of the "Two-armed Bandit" of widespread interest (the "classical" case) is that in which the random variables have binomial distributions with parameters given by:

	X	Y
H_1	p	q
H_2	q	p .

A second example in which the likelihood ratios are distributed alike is furnished here if $p + q = 1$. Hence, for that case, S_1 is the optimal design. Indeed, it is a conjecture that for any choice of p and q , S_1 is optimal; it has been verified to be for $n \leq 8$. Optimal or not, S_1 has the desirable property of being consistent; i.e., Theorem 2.2 holds.

THEOREM 2.2. *Following the design S_1 , the expected value of the average of the first n observations converges to $\max(p, q)$ as $n \rightarrow \infty$.*

PROOF. Assume $p > q$. Then

$$(1) \quad W_1(\zeta, S_1) = \begin{cases} q + (p - q)\zeta & \text{for } \zeta \geq \frac{1}{2}, \\ p - (p - q)\zeta & \text{for } \zeta \leq \frac{1}{2}; \end{cases}$$

and if $\zeta \geq \frac{1}{2}$,

$$(2) \quad \begin{aligned} W_n(\zeta, S_1) &= W_1(\zeta, S_1) + W_{n-1} \left(\frac{p\zeta}{P_T(X=1)} \right) P_T(X=1) \\ &\quad + W_{n-1} \left(\frac{(1-p)\zeta}{P_T(X=0)} \right) P_T(X=0); \end{aligned}$$

while if $\zeta \leq \frac{1}{2}$,

$$(3) \quad \begin{aligned} W_n(\zeta, S_1) &= W_1(\zeta, S_1) + W_{n-1} \left(\frac{q\zeta}{P_T(Y=1)} \right) P_T(Y=1) \\ &\quad + W_{n-1} \left(\frac{(1-q)\zeta}{P_T(Y=0)} \right) P_T(Y=0), \end{aligned}$$

where $P_T(Z=c) = \zeta P_r(Z=c|H_1) + (1-\zeta)P_r(Z=c|H_2)$.

W_1 is clearly convex, symmetric about $\zeta = \frac{1}{2}$, and continuous. So is W_n , since by an inductive argument, W_n is symmetric about $\frac{1}{2}$, (2) and (3) are continuous, and each (by formal differentiation twice) is convex. Also it is easily seen that $W_n(\zeta, S_1) = n[(p-q)\zeta + q]$ for ζ near 1.

Let $a_n(\zeta, S_1) = 1/n W_n(\zeta, S_1)$. Then a_n is convex, continuous, and bounded above by p on $[0, 1]$. Furthermore, $|a'_n(\zeta, S_1)| \leq p - q$. As a consequence of a more general result below, Lemma 3.3, a_n is nondecreasing in n . Hence, $a(\zeta, S_1) = \lim_{n \rightarrow \infty} a_n(\zeta, S_1)$ exists and is convex and continuous on $[0, 1]$. Moreover, since

$na_n(\zeta, S_1)$ satisfies (1) and (2),

$$(4) \quad a(\zeta, S_1) = \begin{cases} a\left(\frac{p\zeta}{P_f(X=1)}\right) P_f(X=1) + a\left(\frac{(1-p)\zeta}{P_f(X=0)}\right) P_f(X=0) & \zeta \geq \frac{1}{2}, \\ a\left(\frac{q\zeta}{P_f(Y=1)}\right) P_f(Y=1) + a\left(\frac{(1-q)\zeta}{P_f(Y=0)}\right) P_f(Y=0) & \zeta \leq \frac{1}{2}. \end{cases}$$

Suppose that the minimum of $a(\zeta, S_1)$ is assumed at $\zeta_0 \geq \frac{1}{2}$. Then it also assumes its minimum at $p\zeta_0/P_f(X=1) > \zeta_0$. By iteration, it assumes its minimum at $(p^n\zeta_0)/[p^n\zeta_0 + q^n(1-\zeta_0)]$, which tends to 1 as $n \rightarrow \infty$. Hence, ζ_0 could be taken to be 1. If, on the other hand, $\zeta_0 < \frac{1}{2}$, the analogous procedure shows that ζ_0 could be taken to be 0. Thus, the minimum of $a(\zeta, S_1)$ is assumed either at 0 or 1. But $a(0, S_1) = a(1, S_1) = p$, which establishes the theorem.

3. A generalized "Two-armed Bandit." This section is concerned with the Bayes problem of maximizing the expected number of successes in n trials when at each trial we are free to choose between two binomial random variables, X and Y , whose probabilities of success, p and q , respectively, are unknown, but a known a priori distribution, $F(p, q)$, is specified.

The special case where $F(p, q)$ concentrates at the two fixed points (p, q) and (q, p) with probabilities ζ and $1 - \zeta$, respectively, leads to the "classical" problem considered in Section 2.2.

Let S denote a strategy for choosing between X and Y and $W_n(p, q, S)$ denote the expected number of successes in following S for n plays for given p and q . Then the expected number of successes is

$$(3.1) \quad W_n(F, S) = \int_0^1 \int_0^1 W_n(p, q, S) dF(p, q).$$

We will find it convenient sometimes to express this as $W_n(dF, S)$. The best strategy is the one maximizing $W_n(F, S)$. Since n is finite, the maximum exists.

Example 1. Suppose $p + q = 1$ with probability 1; i.e., $F(p, q)$ is of the form $F(p, 1-p)$. In this case a success or failure of X is equivalent to (gives the same information as) a failure or success with Y , respectively.

S_1 is optimal in this case; for let $F_{k-1}(p, q)$ denote the a posteriori probability after $(k-1)$ plays and S_{n-k} the optimal strategy for $(n-k)$ plays. Then using X followed by S_{n-k} yields

$$\begin{aligned} \int p dF_{k-1} + W_{n-k}(p dF_{k-1}, S_{n-k}) \int p dF_{k-1} \\ + W_{n-k}((1-p) dF_{k-1}, S_{n-k}) \int (1-p) dF_{k-1}, \end{aligned}$$

* This section represents an extension of some preliminary work by S. Johnson and S. Karlin at The RAND Corporation.

and using Y followed by S_{n-k} yields

$$\int q dF_{k-1} + W_{n-k}(q dF_{k-1}, S_{n-k}) \int q dF_{k-1} \\ + W_{n-k}((1-q) dF_{k-1}, S_{n-k}) \int (1-q) dF_{k-1}.$$

Since $q = 1 - p$, X is the optimal play if and only if $\int p dF_{k-1} \geq \int q dF_{k-1}$; i.e., S_1 is optimal. This example is related to the result of Theorem 2.1, which embraces a special case of $F(p, q) = F(p, 1 - p)$.

3.1. Our first task is to obtain the complete strategy for $n = 2$ when the a priori distribution is $F(p)G(q)$. It is important to notice that, if the number of trials is n , then only designs which are functions of the first n moments, μ_1, \dots, μ_n of F and μ'_1, \dots, μ'_n of G , need be considered. This is a consequence of the fact that the expected yield for any strategy is an expression involving at most these moments. Thus all strategies describing a first move can be expressed in terms of functions $T_j(\mu_1, \dots, \mu_n, \mu'_1, \dots, \mu'_n)$ such that if

$$T_j(\mu_1, \dots, \mu_n, \mu'_1, \dots, \mu'_n) \geq 0,$$

then X is chosen at the first trial; otherwise Y is used first.

Suppose for definiteness that $\mu_1 \geq \mu'_1$; we determine necessary and sufficient conditions that X be used first when $n = 2$. Using the fact that on the last trial one chooses the random variable having greatest expected value, if X is used first the expected yield is

$$(1) \quad \mu_1 + \mu_1 \left(\frac{\mu_2}{\mu_1} \right) + (1 - \mu_1) \max \left\{ \mu'_1, \frac{\mu_1 - \mu_2}{1 - \mu_1} \right\}.$$

Since $(1) \geq 2\mu_1 \geq \mu_1 + \mu'_1 \geq 2\mu'_1$, X followed by optimal is better than Y followed always by X , which is better than Y followed always by Y . Of the other two strategies starting with Y , the one requiring X if $Y = 1$ has expected yield $2\mu'_1 + \mu_1\mu'_1 - \mu'_2$, which can be shown to be less than or equal to that for the strategy requiring Y if $Y = 1$, namely,

$$(2) \quad \mu'_1 + \mu'_2 + (1 - \mu'_1)\mu_1.$$

Upon comparison, $(2) \leq (1)$ if and only if

$$(3.1.1) \quad \text{either } \mu_2 \geq \mu'_2 \text{ or } \mu_1 + \mu_1\mu'_1 \geq \mu'_1 + \mu'_2.$$

Combining and rewriting in a symmetric form, we have

LEMMA 3.1. *If $n = 2$ and p and q are independent with moments μ_i and μ'_i , then X is used on the first trial if and only if*

$$\max\{\mu_2 - \mu'_1\mu_1, \mu_1 - \mu'_1\} \geq \max\{\mu'_2 - \mu_1\mu'_1, \mu'_1 - \mu_1\}.$$

Our next theorem shows that in almost all circumstances of independent a priori distributions for p and q , the optimal design and S_1 cannot agree.

THEOREM 3.1. *If p and q are independent with a priori distributions $F(p) = \int_0^p \phi(t) dt$ and $G(q) = \int_0^q \psi(t) dt$, where ϕ and ψ are continuous and positive for*

$0 < t < 1$, then there exists an n such that for n trials, the optimal design does not agree with S_1 .

PROOF. (By contradiction.) Suppose for definiteness that $1 > \int_0^1 t\phi(t) dt = b \geq a = \int_0^1 t\psi(t) dt$. According to S_1 it is clear that X is used first and, by the Schwarz inequality, that we do not change random variables if a success occurs.

It is easily shown, in view of the hypothesis on ϕ , that if r and s tend to infinity so that $r/(r+s) \rightarrow t_0$, then

$$(1) \quad \frac{\int_0^1 t^{r+1}(1-t)^s \phi(t) dt}{\int_0^1 t^r(1-t)^s \phi(t) dt} \rightarrow t_0.$$

(This can also be obtained as a consequence of the law of large numbers where the relative frequency of success tends to t_0 .) Hence, taking $t_0 > a + \epsilon/2$, we can choose r and s so that

$$(2) \quad \left| \frac{r}{r+s} - a \right| < \epsilon, \quad a + \epsilon > \frac{\int_0^1 t^{r+i+1}(1-t)^s \phi(t) dt}{\int_0^1 t^{r+i}(1-t)^s \phi(t) dt} > a \quad \text{for } i = 0 \text{ and } 1.$$

Furthermore, ϵ may be chosen sufficiently small that also

$$\int t^2 \phi(t) dt > \left[\int t \phi(t) dt \right]^2 + 3\epsilon.$$

Now let $n = r + s + 2$ and suppose that the first r plays resulted in successes with X and the next s plays were failures with X . This agrees with the procedure prescribed by S_1 and has positive probability of occurrence. There are now two plays left and, from (2), S_1 requires X on the next step. However, Lemma 3.1 gives necessary and sufficient conditions that the use of X is optimal and we show that these are violated.

There are two steps left and the a posteriori probability distribution is $F^1(p)G(q)$ where

$$F^1(p) = \frac{p^r(1-p)^s \phi(p)}{\int_0^1 t^r(1-t)^s \phi(t) dt} \quad \text{and} \quad G(q) = \int_0^q \psi(t) dt.$$

On account of (2),

$$\int_0^1 p dF^1(p) > \int_0^1 q \psi(q) dq,$$

and on direct calculation it is seen that both of the inequalities of (3.1.1) are violated. Hence, following S_1 , we arrive at a nonoptimal yield, and the theorem is established.

3.2. S_1 may be described as that procedure requiring at each step the random

variable that would be optimal, were there but one trial remaining. In a similar spirit, let S_j be the strategy which requires at each trial the random variable that would be optimal were there j trials remaining, with the understanding that if fewer than j trials remain, then the optimal procedure is followed. (The strategy S_2 for p and q independent is determined by the relations given in Lemma 3.1.)

We have, thus, a sequence of strategies, S_1, S_2, \dots, S_n . For a series of n trials, S_n is the optimal strategy and hence $W_n(F, S_n) \geq W_n(F, S_j)$ for all $j < n$. Intuitively one might expect that the $W_n(F, S_j)$ are nondecreasing in j ; i.e., the more steps ahead we take into account, the better the strategy. However, it can be shown that there exist a priori distributions such that for $n = 3$, $W_3(F, S_1) > W_3(F, S_2)$. The details are omitted.

3.3. The next principle examined is that of "staying on a winner": Does the optimal strategy have the property that whenever a success occurs, the same random variable is required on the next trial? S_1 , for instance, has this property. However, it is not always a characteristic of an optimal strategy, as the following example shows.

Suppose $F(p, q)$ concentrates probability 0.8 on (0.1, 0) and 0.2 on (0.9, 1). It can be shown that for this example we must stay on a loser but switch from a winner for the case of $n = 2$.

A property related to the intuitive notion of staying on a winner is that of "monotonicity," which we discuss for p and q independent. Let $S^*(n, FG)$ denote the optimal strategy for n trials against a priori FG and let $dF^* = p dF / \int p dF$ and $dF' = (1 - p) dF / \int (1 - p) dF$, with G^* and G' similarly defined. $S^*(n, FG)$ will be termed monotone if:

- (i) $S^*(n, F'G)$ allows X first $\Rightarrow S^*(n, F^*G)$ requires X first,
- (ii) $S^*(n, FG')$ allows Y first $\Rightarrow S^*(n, FG^*)$ requires Y first,

and

- (iii) $S^*(n, F^*G)$ allows Y first $\Rightarrow S^*(n, FG^*)$ requires Y first.

For instance, (i) is to be thought of as: if a prior "free" observation of X were allowed, then, if we might use X on the first trial, even if X had failed on the prior trial, we should certainly use X on the first trial, if the prior trial had resulted in a success with X .

LEMMA 3.2. If for $1 \leq k \leq n - 1$ and for all F and G , $S^*(k, FG)$ is monotone, then $S^*(n, FG)$ stays on a winner.

The proof is omitted.

With the aid of the results of Lemma 3.1, it can be verified directly that, for $n = 1$ and $n = 2$, the optimal strategy is monotone and, therefore, for $n = 3$ and p and q independent, the optimal strategy stays on a winner.

The general monotonicity property for independent parameters remains an open question.

3.4. In using any strategy, S , let $Z_r = 1$ if the random variable used on the r -th trial wins, and $Z_r = 0$ otherwise; i.e., Z_r is the contribution of the r -th trial. The last property considered is whether, for S^* denoting the optimal S ,

$E[Z_r | S^*]$ is monotone increasing in r . We show first that $E[Z_r | S_1]$ is non-decreasing in r , and second that $E[Z_r | S^*]$ may decrease.

LEMMA 3.3. $E[Z_r | S_1]$ is nondecreasing in r for every initial distribution F .

PROOF. It is enough to prove the result for the first two trials. Suppose $Z_1 = X$, then $E[Z_1 | S_1] = \int p dF$. But $E[Z_2 | S_1] = E[E[Z_2 | Z_1, S_1]] \geq E[E[X | Z_1, S_1]] = E[Z_1 | S_1]$.

In contrast to this result, consider the case of $n = 3$,

$$F(p) = p^5(1-p)^2 / \int p^5(1-p)^2 dp,$$

and $G(q) = q$. For optimal return, X should be employed first with expected return from the first trial of 0.6. If success results, then X is used again, while if failure occurs, then the criteria of Lemma 3.1 require Y . The a priori expected yield from the second trial is $64/110 < 66/110 = 0.6$.

4. The case of one known and one random probability of success. In this section we examine in detail the situation in which X has a binomial distribution with $p = \Pr(X = 1)$ unknown but selected by a known a priori distribution, F , while Y has a binomial distribution with known parameter, q .

The results of the preceding section were informative largely in a negative sense; there are many nice properties which optimal strategies do not possess. Many properties which seemed obvious but which were not in general enjoyed by optimal strategies in the general case, are held by the optimal strategy when one of the random variables has a known distribution. Hence, the rather detailed proofs in this section.

We establish a series of lemmas describing some properties and the form of the optimal strategy and then obtain an explicit statement of it.

LEMMA 4.1. If Y is required at any trial according to an optimal strategy, then Y is required thereafter.

PROOF. First, it is easily seen that if at any trial Y is required, then the optimal choice for the next trial is independent of whether Y wins or not.

Now suppose that at some trial, let us say at the first one, Y is required and is used r times, but that X is allowed on the $(r+1)$ -st trial. Then the expected winnings are

$$(1) \quad rq + W_{n-r-1}(F^*, q) \int_0^1 p dF + W_{n-r-1}(F', q) \int_0^1 (1-p) dF,$$

where F^* and F' are the a posteriori probabilities defined in Section 3.3, and $W_k(F, q)$ is the expected gain against a priori F in k trials pursuing an optimal strategy. But using X first followed by r trials of Y yields the same amount, contradicting the fact that Y was required on the first trial. Hence Y must be required throughout.

As a consequence of Lemma 4.1, we can characterize an optimal strategy. We use the notation

$$\mu_i = \int_0^1 p^i dF.$$

LEMMA 4.2. *There exists a function, Q , of n and F such that for n trials remaining and F the a priori distribution of p at that time, Y is required if and only if $q > Q(n, F)$.*

PROOF. From Lemma 4.1, Y is required if and only if

$$(1) \quad nq > \mu_1 + \mu_1 W_{n-1}(F^s, q) + (1 - \mu_1)W_{n-1}(F^f, q) = nK_n(F, q).$$

Now $W_1(F, q) = \max \{q, \mu_1\}$ and hence is nondecreasing convex in q for all F . By easy induction, $W_n(F, q)$ is nondecreasing convex in q for all F and n and, hence, so is $K_n(F, q)$. Since $K_n(F, 0) = \mu_1 > 0$ and $K_n(F, 1) = 1 - (1 - \mu_1)/n < 1$, it follows that for each n and F there is a point $Q(n, F)$ such that $q > Q(n, F)$ if and only if $q > K_n(F, q)$ —i.e., if and only if Y is required.

We shall adopt the convention that if $q = Q(n, F)$ we shall always use X , giving us a definite optimal strategy:

If $q > Q(n, F)$, use Y for all n trials. If $q \leq Q(n, F)$, use X on the first trial and compute the a posteriori distribution of p , F' , and compare q and $Q(n-1, F')$, following the above rules for choice at the second trial, etc.

Having characterized the optimal strategy, we turn to a series of lemmas describing more precisely its form and properties.

LEMMA 4.3. *For all F and $n \geq 2$, $Q(n, F) \geq Q(n-1, F)$.*

PROOF. Suppose the contrary. Then for $Q(n, F) < q < Q(n-1, F)$, Y would be required on the first trial and X on the second, contradicting Lemma 4.1.

LEMMA 4.4. *For all F , q , and n ,*

$$W_n(F^s, q) \geq W_n(F, q) \geq W_n(F^f, q).$$

PROOF.

$$I: q \geq \max\{Q(n, F^f), Q(n, F), Q(n, F^s)\}.$$

Then $nq = W_n(F^s, q) = W_n(F, q) = W_n(F^f, q)$.

$$II: q \leq \min\{Q(n, F^f), Q(n, F), Q(n, F^s)\}.$$

We proceed by induction. The lemma holds for $n = 1$, since for all q and F ,

$$(1) \quad \max\left\{q, \frac{\mu_2}{\mu_1}\right\} \geq \max\{q, \mu_1\} \geq \max\left\{q, \frac{\mu_1 - \mu_2}{1 - \mu_1}\right\}.$$

In the case under consideration,

$$(2) \quad W_n(F^s, q) = \frac{\mu_2}{\mu_1} + \frac{\mu_2}{\mu_1} W_{n-1}(F^{ss}, q) + \left(1 - \frac{\mu_2}{\mu_1}\right) W_{n-1}(F^{sf}, q) = \frac{\mu_2}{\mu_1} + A_n,$$

$$(3) \quad W_n(F, q) = \mu_1 + \mu_1 W_{n-1}(F^s, q) + (1 - \mu_1)W_{n-1}(F^f, q) = \mu_1 + B_n,$$

$$(4) \quad W_n(F^f, q) = \frac{\mu_1 - \mu_2}{1 - \mu_1} + \frac{\mu_1 - \mu_2}{1 - \mu_1} W_{n-1}(F^{fs}, q) + \left(1 - \frac{\mu_1 - \mu_2}{1 - \mu_1}\right) W_{n-1}(F^{ff}, q) = \frac{\mu_1 - \mu_2}{1 - \mu_1} + C_n.$$

By the induction hypothesis,

$$(5) \quad W_{n-1}(F^{ns}, q) \geq W_{n-1}(F^s, q) \geq W_{n-1}(F^{sf}, q) \\ = W_{n-1}(F^{fs}, q) \geq W_{n-1}(F^f, q) \geq W_{n-1}(F^{ff}, q);$$

and it is easily shown that since

$$\frac{\mu_2}{\mu_1} > \mu_1 > \frac{\mu_1 - \mu_2}{1 - \mu_1},$$

$A_n \geq B_n \geq C_n$. Thus the lemma is established for this case.

As a consequence of Case II, $Q(n, F^f) \leq \min\{Q(n, F), Q(n, F^s)\}$. For if (say) $Q(n, F) = \min\{Q(n, F), Q(n, F^s)\} < Q(n, F^f)$, then for $q = Q(n, F)$, $nq = W_n(F, q) < W_n(F^f, q)$, a contradiction of the case just established.

$$\text{III: } Q(n, F^f) < q \leq \min\{Q(n, F), Q(n, F^s)\}.$$

Then $W_n(F^f, q) = nq \leq \min\{W_n(F, q), W_n(F^s, q)\}$. But by induction argument parallel to that for Case II, it is shown that $W_n(F, q) \leq W_n(F^s, q)$.

From Case III it follows by the same reasoning as above that $Q(n, F) \leq Q(n, F^s)$. Hence, there is only one remaining case.

$$\text{IV: } Q(n, F^f) \leq Q(n, F) < q < Q(n, F^s).$$

Immediately, $W_n(F^f, q) = W_n(F, q) = nq \leq W_n(F^s, q)$, and the lemma is established.

Interspersed in the proof just completed is the proof of

LEMMA 4.5.

$$Q(n, F^f) \leq Q(n, F) \leq Q(n, F^s).$$

LEMMA 4.6. *Following the optimal strategy, if a success occurs on any trial, then the same random variable is used on the next trial—i.e., stay with a winner.*

PROOF. In view of Lemma 4.1, we need only show that if X is required and wins, then X is required on the next trial. It is clearly sufficient to show this for the first trial. Suppose to the contrary that $Q(n, F) \geq q > Q(n-1, F^s)$. By Lemma 4.3, $Q(n-1, F^s) \geq Q(n-2, F^s) \geq \dots \geq Q(1, F^s)$ and, clearly, $Q(1, F^s) \geq \mu_2/\mu_1 > \mu_1$. Hence, $q > \mu_1$.

By Lemma 4.5, $q > Q(n-1, F^f)$ also. Consequently Y is required on the second trial, regardless of the outcome of first. Then

$$(1) \quad nq \leq W_n(F, q) = \mu_1 + (n-1)q.$$

Hence, $q \leq \mu_1$, and we have a contradiction.

LEMMA 4.7. *The a priori expected value of the yield on the r -th step is nondecreasing in r when using an optimal strategy.*

PROOF. The proof can be obtained with the aid of the foregoing lemmas; it is left as an exercise for the reader.

As we have noted in Section 3, Lemma 4.7 is not true, in general, while Lemma 4.6. is the "stay on a winner" rule which, appealing as it is, does not hold in general.

With the above lemmas we are in a position to determine explicitly the value of $Q(n, F)$. Assume that $q = Q(n, F)$; then the optimal strategy has the following form for appropriate k_i .

(4.1) (A) Observe X until a failure occurs.

(B) There exists an integer $k_1 \geq 0$ such that if at least k_1 successes preceded the first failure, continue with X ; otherwise switch to Y for the remaining trials.

(C) There is an integer $k_2 \geq 0$ attached to the second failure such that if at least $k_1 + k_2$ successes with X precede the second failure of X , continue with X ; otherwise switch to Y for the remaining trials.

(D) In general, let S_r be the number of successes that precede the r -th failure of X . If $S_r \geq k_1 + k_2 + \dots + k_r$, continue with X ; otherwise switch to Y for the remaining trial.

Thus, any sequence $k = (k_1, k_2, \dots, k_n)$ of integers, $0 \leq k_i \leq n$, corresponds to a strategy of the same form as the optimal.

Let E_k denote expectation given k and F , and E_{kp} denote expected value given k and p .³ In using any strategy for n trials, X will be used a certain number, N_x , of times, and there will be a certain number, S_x , of successes with X ; similarly for Y .

THEOREM 4.1.

$$Q(n, F) = \max_k \left\{ \frac{E_k[S_x]}{E_k[N_x]} \right\}.$$

PROOF. $q = Q(n, F)$ implies $nq = W_n(F, q)$. But since the optimal strategy corresponds to a sequence k , this is equivalent to $nq = \max_k \{E_k[S_x] + E_k[S_y]\}$. However, $E_k[S_y] = qE_k[N_y]$, and neither $E_k[S_x]$ nor $E_k[N_y]$ depends on q . Hence $q = Q(n, F)$ implies each of the following equivalent statements: $nq = \max_k \{E_k[S_x] + qE_k[N_y]\} \Leftrightarrow q \geq E_k[S_x]/E_k[N_x]$ for all k with equality for some $\Leftrightarrow q = \max_k E_k[S_x]/E_k[N_x]$.

COROLLARY.

$$Q(n, F) = \max_k \left\{ \frac{\int_0^1 E_{kp}[N_x] p \, dF}{\int_0^1 E_{kp}[N_x] \, dF} \right\} = \max_k \left\{ \frac{\int_0^1 (n - E_{kp}[N_y]) p \, dF}{\int_0^1 (n - E_{kp}[N_y]) \, dF} \right\}.$$

³ The authors are indebted to the referee for suggesting the following derivation of $Q(n, F)$, which is somewhat simpler and more illuminating than that originally used to obtain the result.

We give two methods of evaluating $Q(n, F)$. The first proceeds by obtaining directly a formula for $E_{kp}[N_\nu]$ and yields

$$(4.2) \quad E_{kp}[N_\nu] = \sum_{j=1}^n (n-j)p^{j-\phi(j)-1}(1-p)^{\phi(j)+1} \cdot \sum^* \prod_{i=0}^{\phi(j)-1} \binom{k_{i+1} + f_i - 1}{f_i} \binom{j - \sum_{r=1}^{\phi(j)} k_r - \sum_{r=1}^{\phi(j)-1} f_r - 1}{\phi(j) - \sum_{r=1}^{\phi(j)-1} f_r},$$

where $\phi(j) = \max\{i: \sum_{r=1}^i k_r + i < j\}$ = number of failures of X in the first $(j-1)$ trials and \sum^* denotes the sum over all choices of f_i 's such that $f_i \geq 0$; $f_i = 0$, if $k_{i+1} = 0$, $f_0 = 0$; $\sum_{i=0}^{\phi(j)} f_i \leq \nu$ and $\sum f_i = \phi(j)$ (f_i denotes the number of failures between the $(k_1 + k_2 + \dots + k_i)$ -th and $(k_1 + k_2 + \dots + k_{i+1})$ -th success).

The second proceeds by obtaining directly a formula for $E_{kp}[S_z] = E_{kp}[N_z]p$. While more complicated in appearance and derivation, it is the result of a direct counting.

$$(4.3) \quad E_{kp}[S_z] = \sum_{r=0}^{z-n} I_r,$$

where

$$\begin{aligned} I_r = & \sum_{a_1=0}^1 \binom{a_1-1}{a_1-1} + \dots \\ & + \binom{k_2 + a_1 - 1}{a_1 - 1} \sum_{a_2=0}^{2-a_1} \left\{ \binom{a_2-1}{a_2-1} + \dots + \binom{k_3 + a_2 - 1}{a_2 - 1} \right\} \\ & + \dots + \sum_{a_r=0}^{r-a_1-a_2-\dots-a_{r-1}} \left\{ \binom{a_r-1}{a_r-1} + \dots + \binom{k_{r+1} + a_r - 1}{a_r - 1} \right\} \\ & \cdot \left[\binom{b_r}{b_r} p + \binom{b_r+1}{b_r} p^r + \dots + \binom{b_r + n - \sum_{i=1}^r k_i - r - 1}{b_r} \right. \\ & \left. p^{n - \sum_{i=1}^r k_i - r} \right] p^{k_1+k_2+\dots+k_r} (1-p)^r, \end{aligned}$$

with $b_r = r - a_1 - a_2 - \dots - a_r$, and we interpret $\binom{-1}{-1} = 1$ and $\binom{c}{-1} = 0$, for $c \neq -1$.

Some special cases are worth noting:

$$Q(2, F) = \frac{\int_0^1 (p + p^2) dF}{\int_0^1 (1 + p) dF},$$

$$Q(3, F) = \max \left\{ \frac{\int_0^1 (p + p^2 + p^3) dF}{\int_0^1 (1 + p + p^2) dF}, \frac{\int_0^1 (p + 2p^2) dF}{\int_0^1 (1 + 2p) dF} \right\},$$

Each term of $Q(3, F)$ can occur; e.g., for $F(p) = p$, the first is the maximum (value 13/22), while for $F(p) = p^{1/6}$, the second is the maximum (value 23/88).

The expression for $Q(n, F)$ cannot be simplified in any essential way, which again testifies to the complex nature of the optimal strategies in sequential design problems.

If one chooses $k_1 = r$, $k_2 = k_3 = \dots = 0$, then

$$(4.4) \quad \frac{E_k[S_x]}{E_k[N_x]} = \frac{\mu_1 + \mu_2 + \dots + \mu_r + (n - r)\mu_{r+1}}{1 + \mu_1 + \dots + \mu_{r-1} + (n - r)\mu_r},$$

where $\mu_i = \int_0^1 p^i dF(p)$. For distributions such that $\sum_{r=1}^{\infty} \mu_r = +\infty$, at least, a reasonable approximation to $Q(n, F)$ may be had by taking $r = n - 1$ and using

$$L(n, F) = \frac{\mu_1 + \mu_2 + \dots + \mu_n}{1 + \mu_1 + \dots + \mu_{n-1}}$$

in place of $Q(n, F)$. For the uniform distribution, Q and L coincide for $n \leq 4$, but not for larger n . It is worth remarking that L shares many of the properties of Q .

LEMMA 4.8. L is nondecreasing in n , and $L(n, F') \geq L(n, F) \geq L(n, F')$.

The proof is omitted.

We close by noting that if the number of trials is sufficiently large, one should almost always commence by using X . More precisely, we see from (4.4) that as $n \rightarrow \infty$, $Q(n, F)$ becomes at least $(\mu_{r+1})/\mu_r$, and this for every r . But $(\mu_{r+1})/\mu_r$ is the expected value of p given r successes which will tend to the supremum of the spectrum of F as r increases. Clearly, if q is greater than the supremum of the spectrum of F , one would never play X ; while if q is less than the supremum, for all sufficiently large n , X should be used first. Finally, for a fixed n we have the following intuitive result:

LEMMA 4.9. *Given n and F , Y should never be used if*

$$q \leq \frac{\int_0^1 p(1-p)^{n-1} dF}{\int_0^1 (1-p)^{n-1} dF}.$$

The rigorous proof may be supplied by the reader.

5. An applied problem. An interesting problem in industrial inspection is closely related to the problem of Section 4. Suppose that lots of n items are produced by a process having probability p of producing a defective where p varies from lot to lot according to an a priori distribution, $F(p)$. Let the loss per defective item accepted be unity and the cost of inspection be c per item inspected ($c < 1$). Items are drawn and inspected (defective items found being replaced by good items at no additional cost) until a sequential stopping rule terminates inspection, at which point the remainder of the lot is accepted. A stopping rule is desired which will minimize the expected loss.

One may proceed to attack this problem in the spirit of Section 4 and find a completely analogous series of lemmas, culminating in the theorem that for n items remaining and a priori distribution F , it is optimal to inspect another item or to accept the remaining n according as c is less than or greater than $Q(n, F)$. The same result is more immediately obtained by noting that the problem is equivalent to finding a rule to maximize the gain if one wins c for each item not inspected, nothing for each good item inspected, and one for each defective item inspected (and replaced). This latter problem is precisely that treated in Section 4.

REFERENCES

- [1] THOMPSON, W. R., "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, Vol. 25 (1933), pp. 285-294.
- [2] ROBBINS, H., "Some aspects of the sequential design of Experiments," *Bull. Amer. Math. Soc.*, Vol. 58, No. 5 (1952), pp. 527-535.

TABLES FOR COMPUTING BIVARIATE NORMAL PROBABILITIES

By DONALD B. OWEN

Sandia Corporation

1. Introduction. Various tables have been published for obtaining probabilities over rectangles for correlated bivariate normal variables. Some of these tables give the probabilities as functions of three parameters (see [1], [2], and [3]). Others tabulate related two-parameter families from which these probabilities may be computed (see [3], [4], [5], [6], and [7]). The tables given here are of the latter type. They have been computed for use with a special two-dimensional interpolation scheme, which is described in Section 4. These new tabulations reduce considerably the amount of interpolation work required over that needed with previous tables. The function tabulated also eliminates an arctangent function from the formula for the bivariate normal over a region outside of a rectangle as compared with the formula for Nicholson's tabulation in [5]. Section 3 contains a derivation of the formulas given in Section 2 for using a two-parameter table to compute probabilities over rectangles. The tables given below should prove very useful, since examples where bivariate normal integrals over polygons are needed to solve practical problems abound in the literature. For example, see [6], [8], [9], and [10].

The usefulness of the $T(h, a)$ function tabulated below was also recognized by Professor Harry A. Bender, University of Rhode Island, who submitted, after this paper was received by the editor, a somewhat shorter tabulation than given here. An abstract of Professor Bender's paper appears in [15].

For h and $a > 0$, $T(h, a)$, the function tabulated, gives the volume of an uncorrelated bivariate normal distribution with zero means and unit variances over the area between $y = ax$ and $y = 0$ and to the right of $x = h$, i.e., the area shaded in Fig. 1.

Cadwell in [11] gives a method for obtaining the volume of a bivariate normal over any polygon. In Fig. 2, if AB is a side of any polygon, then the volume over the shaded area for an uncorrelated bivariate normal with zero means and unit variances is given by

$$T(h, a_2) - T(h, a_1)$$

for $a_2 > a_1$, where h is the length of the perpendicular from the origin to the line through AB and a_1h is the distance from the foot of the perpendicular, C , to B and a_2h is CA . If C lies between A and B , then the T -functions are added instead of subtracted. By composition of volumes like this, it is possible to obtain the volume over the area outside of any polygon. Section 2 includes some useful formulas for doing this.

Received December 8, 1955; revised April 2, 1956.

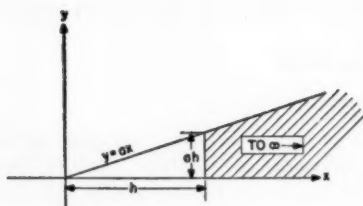


FIG. 1

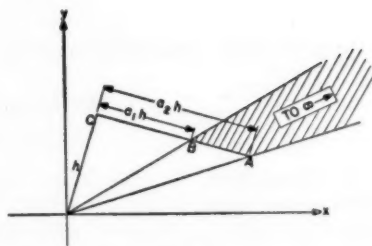


FIG. 2

FIG. 1. The area over which $T(h, a)$ gives the volume of a standardized bivariate normal with correlation zero.

FIG. 2. A typical area for computing the bivariate normal over a polygon.

2. Summary of formulas. The fundamental formula for finding volumes over rectangles is

$$(2.1) \quad B(h, k; \rho) = \frac{1}{2}G(h) + \frac{1}{2}G(k) - T(h, a_h) - T(k, a_k) - \begin{cases} 0 \\ \frac{1}{2} \end{cases},$$

where the upper choice is made if $hk > 0$ or if $hk = 0$ but $h + k \geq 0$, and the lower choice is made otherwise, where

$$(2.2) \quad a_h = \frac{k}{h\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}}, \quad a_k = \frac{h}{k\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}},$$

and where $B(h, k; \rho)$ is the volume of a bivariate normal with zero means and unit variances and correlation ρ over the lower left-hand quadrant of the xy -plane when divided at $x = h$ and $y = k$, $G(h)$ is the univariate normal with zero mean and unit variance integral from minus infinity to h , and $T(h, a)$ is the function tabulated below.

The T -function is tabulated only for $0 < a \leq 1$, and ∞ , but it is possible to obtain values for $1 < a < \infty$ by use of the following formula:

$$(2.3) \quad T(h, a) = \frac{1}{2}G(h) + \frac{1}{2}G(ah) - G(h)G(ah) - T\left(ah, \frac{1}{a}\right).$$

Values for negative a or h may be obtained by using

$$(2.4) \quad T(h, -a) = -T(h, a)$$

and

$$(2.5) \quad T(-h, a) = T(h, a).$$

Note that (2.3) requires a to be positive and hence when a is negative, first apply (2.4) and then (2.3).

Other useful formulas are:

$$T(h, 0) = 0,$$

$$T(0, a) = \frac{1}{2\pi} \arctan a,$$

$$T(h, 1) = \frac{1}{2}G(h)[1 - G(h)],$$

and

$$T(h, \infty) = \begin{cases} \frac{1}{2}[1 - G(h)] & \text{if } h \geq 0, \\ \frac{1}{2}G(h) & \text{if } h \leq 0. \end{cases}$$

For finding volumes of the general correlated bivariate normal over polygons, the first step is to make a rotation and stretching of the axes to reduce the function under the integral to the form of the T -function. A transformation that will do this is

$$u = \frac{1}{\sqrt{2 + 2\rho}} \left[\frac{x - \mu_x}{\sigma_x} + \frac{y - \mu_y}{\sigma_y} \right],$$

$$v = \frac{-1}{\sqrt{2 - 2\rho}} \left[\frac{x - \mu_x}{\sigma_x} - \frac{y - \mu_y}{\sigma_y} \right],$$

for $\rho^2 < 1$, where μ_x , μ_y are the means of the X and Y variables and σ_x , σ_y are the standard deviations of the X and Y variables, respectively. This will take the original polygon into another polygon in the uv plane. The vertices of the new polygon should be computed and a graph drawn. For each side of the polygon the volume over a region like that shown in Fig. 2 may be computed with the aid of these formulas:

$$h = \frac{|h_1 k_2 - h_2 k_1|}{\sqrt{(h_2 - h_1)^2 + (k_2 - k_1)^2}},$$

$$a_1 = \frac{|h_1(h_2 - h_1) + k_1(k_2 - k_1)|}{|h_1 k_2 - h_2 k_1|},$$

$$a_2 = \frac{|h_2(h_2 - h_1) + k_2(k_2 - k_1)|}{|h_1 k_2 - h_2 k_1|},$$

where the vertical bars indicate absolute value and where (h_1, k_1) and (h_2, k_2) are the coordinates of two adjacent vertices on the polygon. With the aid of the graph, these volumes are then easily combined to give the volume over the outside (or inside) of the polygon.

3. Derivation of the relationship between the bivariate normal and the tabulated function.

Let

$$(3.1) \quad B(h, k; \rho) = \frac{1}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^h \int_{-\infty}^k \exp[-\frac{1}{2}(x^2 - 2\rho xy + y^2)/(1-\rho^2)] dx dy,$$

$$(3.2) \quad G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{1}{2}t^2) dt,$$

and

$$(3.3) \quad T(h, a) = \frac{1}{2\pi} \int_0^a \frac{\exp[-\frac{1}{2}h^2(1+x^2)]}{1+x^2} dx.$$

It is also convenient to have a second form of (3.3), which is the function in Tables A, B and C. It may be obtained by differentiating with respect to h and then reintegrating. The result is

$$(3.4) \quad T(h, a) = \frac{-1}{2\pi} \int_0^h \int_0^{az} \exp[-\frac{1}{2}(x^2 + y^2)] dy dx + \frac{\arctan a}{2\pi}.$$

The T -function is related to the V -function tabulated by Nicholson in [5] as follows:

$$T(h, a) = \frac{1}{2\pi} \arctan a - V(h, ah).$$

If (3.4) is integrated by parts,

$$(3.5) \quad T(h, a) = \begin{cases} \frac{1}{2}G(h) + \frac{1}{2}G(ah) - G(h)G(ah) - T(ah, 1/a) & \text{if } a \geq 0, \\ \frac{1}{2}G(h) + \frac{1}{2}G(ah) - G(h)G(ah) - T(ah, 1/a) - \frac{1}{4} & \text{if } a < 0. \end{cases}$$

It will be shown that (3.1) can be expressed as a function of expressions like (3.2) and (3.3). If (3.1) is differentiated with respect to ρ , then integration with respect to x and y can be effected. Integrating that result with respect to ρ yields

$$(3.6) \quad B(h, k; \rho) = \frac{1}{2\pi} \int_0^\rho (1-z^2)^{-1/2} \exp[-\frac{1}{2}(h^2 - 2h kz + k^2)/(1-z^2)] dz + G(h)G(k).$$

From this $B(0, 0; \rho) = 1/(2\pi) \arcsin \rho + \frac{1}{4}$, a well-known result (see [12], [13], and [14]). Now (3.6) may be rewritten as

$$\begin{aligned} B(h, k; \rho) &= \frac{1}{2\pi} \int_0^\rho (1-z^2)^{-1/2} \frac{h(h-kz)}{h^2 - 2h kz + k^2} \exp[-\frac{1}{2}(h^2 - 2h kz + k^2)/(1-z^2)] dz \\ &\quad + \frac{1}{2\pi} \int_0^\rho (1-z^2)^{-1/2} \frac{k(k-hz)}{h^2 - 2h kz + k^2} \exp[-\frac{1}{2}(h^2 - 2h kz + k^2)/(1-z^2)] dz \\ &\quad + G(h)G(k). \end{aligned}$$

In the integrals above, making the substitutions

$$u = \frac{k - hz}{h\sqrt{1-z^2}} \quad \text{and} \quad v = \frac{h - kz}{k\sqrt{1-z^2}},$$

respectively, produces

$$(3.7) \quad B(h, k; \rho) = T\left(h, \frac{k}{h}\right) + T\left(k, \frac{h}{k}\right) - T\left(h, \frac{k - \rho h}{h\sqrt{1-\rho^2}}\right) \\ - T\left(k, \frac{h - \rho k}{k\sqrt{1-\rho^2}}\right) + G(h)G(k).$$

Applying (3.5) to (3.7), gives

$$(3.8) \quad B(h, k; \rho) = \begin{cases} \frac{1}{2}G(h) - T\left(h, \frac{k - \rho h}{h\sqrt{1-\rho^2}}\right) + \frac{1}{2}G(k) - T\left(k, \frac{h - \rho k}{k\sqrt{1-\rho^2}}\right), \\ \quad \text{if } hk > 0 \text{ or if } hk = 0, h \text{ or } k \geq 0 \\ \frac{1}{2}G(h) - T\left(h, \frac{k - \rho h}{h\sqrt{1-\rho^2}}\right) + \frac{1}{2}G(k) \\ \quad - T\left(k, \frac{h - \rho k}{k\sqrt{1-\rho^2}}\right) - \frac{1}{2}, \text{ if } hk < 0 \text{ or if } hk = 0, h \text{ or } k < 0 \end{cases}$$

which expresses the bivariate normal in terms of the G - and T -functions in a compact form.

A series expression for $T(h, a)$ may be obtained by expanding the numerator of the integrand of (3.3) in the usual exponential series, dividing by the denominator, and integrating term by term. Rearrangement of the terms of this series gives

$$(3.9) \quad T(h, a) = \frac{\arctan a}{2\pi} - \frac{1}{2\pi} \sum_{j=0}^{\infty} c_j a^{2j+1},$$

where

$$c_j = (-1)^j \frac{1}{2j+1} \left[1 - \exp\left(-\frac{1}{2}h^2\right) \sum_{i=0}^j \frac{h^{2i}}{2^{i+1}i!} \right],$$

which converges rapidly for small values of a and h .

The values of $T(h, a)$ given in Tables A, B, and C were computed using the series (3.9). They were checked by using Gauss' seven-point integration formula on (3.3). The tables were also checked by taking differences. These checks show that at the points of tabulation the table is accurate to as many places as given, i.e., to six decimal places.

4. Interpolation in the tables. Table A has a coarse interval in the parameter a and an interval fine enough for ordinary linear interpolation in the parameter h . Table B has intervals in parameter a fine enough for ordinary linear interpolation and has parameter h at a coarse interval. Ordinary linear interpolation

TABLE A OF $\pi(h, a)$
Note that $\pi(h, 0) = 0$

$\frac{a}{h}$.25	.50	.75	1.00	$\frac{a}{h}$.25	.50	.75	1.00
0.00	.038990	.073792	.102416	.125000	0.30	.037240	.070297	.097186	.119018
0.01	.988	.788	.610	.410	0.31	.124	.066	.096811	.117592
0.02	.982	.776	.592	.392	0.32	.005-	.069828	.096811	.123
0.03	.972	.756	.563	.363	0.33	.036882	.584	.122	.116411
0.04	.958	.728	.521	.321	0.34	.756	.333	.095748	.116
0.05	.940	.692	.467	.267	0.35	.627	.076	.365-	.115639
0.06	.918	.649	.424	.202	0.36	.195-	.068812	.094971	.119
0.07	.892	.597	.374	.141	0.37	.359	.512	.111587	.114587
0.08	.862	.538	.335	.092	0.38	.220	.265+	.157	.044
0.09	.829	.470	.284	.357	0.39	.078	.067983	.093736	.113489
0.10	.791	.395-	.821	.207	0.40	.035913	.694	.306	.112922
0.11	.750	.312	.697	.041	0.41	.785-	.399	.092868	.114
0.12	.704	.221	.561	.123860	0.42	.634	.098	.120	.111755+
0.13	.655-	.122	.413	.663	0.43	.179	.066791	.091965-	.135+
0.14	.602	.016	.253	.450+	0.44	.322	.479	.501	.110545-
0.15	.545-	.072902	.082	.223	0.45	.162	.161	.028	.109924
0.16	.484	.780	.109900	.122960	0.46	.034999	.065837	.090548	.108652
0.17	.419	.651	.706	.722	0.47	.834	.508	.060	.108652
0.18	.350+	.514	.501	.449	0.48	.665+	.173	.089564	.107341
0.19	.278	.369	.265-	.162	0.49	.494	.064834	.060	.107341
0.20	.202	.217	.057	.121859	0.50	.320	.189	.088549	.106671
0.21	.122	.058	.099818	.542	0.51	.144	.139	.031	.105993
0.22	.038	.071891	.569	.210	0.52	.03985-	.063784	.087506	.105993
0.23	.037951	.717	.308	.120864	0.53	.783	.124	.104609	.104609
0.24	.860	.535+	.037	.503	0.54	.599	.059	.135-	.103905+
0.25	.766	.347	.098755-	.129	0.55	.113	.062690	.085589	.103
0.26	.668	.151	.462	.119740	0.56	.224	.316	.337	.102473
0.27	.566	.070948	.158	.337	0.57	.033	.061938	.081779	.101745+
0.28	.461	.738	.097844	.118921	0.58	.215	.555+	.010	.101745+
0.29	.352	.521	.520	.492	0.59	.645-	.168	.083645-	.100268
0.30	.240	.297	.186	.048	0.60	.147	.060778	.069	.099519

TABLE A OF $T(h, a)$

$\frac{h}{h}$.25	.50	.75	1.00
0.60	.032117	.060778	.083069	.099519
0.61	.031717	.059814	.081901	.098347
0.62	.031317	.058850	.080712	.097175
0.63	.030917	.057886	.079504	.095996
0.64	.030517	.056922	.078288	.094817
0.65	.030117	.055958	.077073	.093638
0.66	.029717	.054994	.075858	.092459
0.67	.029317	.054030	.074643	.091280
0.68	.028917	.053066	.073428	.090101
0.69	.028517	.052102	.072213	.088922
0.70	.028117	.051138	.070998	.087743
0.71	.027717	.050174	.069783	.086564
0.72	.027317	.049210	.068568	.085385
0.73	.026917	.048246	.067353	.084206
0.74	.026517	.047282	.066138	.083027
0.75	.026117	.046318	.064923	.081848
0.76	.025717	.045354	.063708	.080669
0.77	.025317	.044390	.062493	.079490
0.78	.024917	.043426	.061278	.078311
0.79	.024517	.042462	.060063	.077132
0.80	.024117	.041498	.058848	.075953
0.81	.023717	.040534	.057633	.074774
0.82	.023317	.039570	.056418	.073595
0.83	.022917	.038606	.055203	.072416
0.84	.022517	.037642	.053988	.071237
0.85	.022117	.036678	.052773	.070058
0.86	.021717	.035714	.051558	.068879
0.87	.021317	.034750	.050343	.067699
0.88	.020917	.033786	.049128	.066520
0.89	.020517	.032822	.047913	.065341
0.90	.020117	.031858	.046698	.064162
0.91	.019717	.030894	.045483	.062983
0.92	.019317	.029930	.044268	.061804
0.93	.018917	.028966	.043053	.060625
0.94	.018517	.028002	.041838	.059446
0.95	.018117	.027038	.040623	.058267
0.96	.017717	.026074	.039408	.057088
0.97	.017317	.025110	.038193	.055909
0.98	.016917	.024146	.036978	.054730
0.99	.016517	.023182	.035763	.053551
1.00	.016117	.022218	.034548	.052372

TABLE A OF $\pi(h, a)$

$\frac{a}{h}$.25	.50	.75	1.00	$\frac{a}{h}$.25	.50	.75	1.00
1.20	.018702	.033893	.045317	.050911	1.50	.012372	.022006	.028029	.031172
1.21	.173	.556	.013925+	.169	1.51	.184	.021654	.027553	.030631
1.22	.246	120	.317	.019130	1.52	.011997	.305+	.082	.063
1.23	.019	.032687	.012712	.018696	1.53	.012	.020959	.026616	.029519
1.24	.017794	.256	.112	.017967	1.54	.628	.617	.155+	.028982
1.25	.569	.031828	.041515-	.244	1.55	.446	.279	.025700	.451
1.26	.345+	.462	.040922	.046527	1.56	.266	.019913	.027927	.027927
1.27	.123	.030978	.333	.045315-	1.57	.088	.611	.021801	.410
1.28	.016902	.556	.039703	.109	1.58	.010911	.283	.364	.026889
1.29	.682	.138	.167	.044409	1.59	.736	.018958	.023929	.395+
1.30	.463	.029721	.038590	.043715+	1.60	.563	.637	.500-	.025888
1.31	.205-	.308	.018	.027	1.61	.391	.319	.075+	.468
1.32	.028	.028897	.037450+	.042345+	1.62	.221	.005-	.022656	.021924
1.33	.015813	.488	.036106	.041670	1.63	.053	.017694	.242	.447
1.34	.599	.083	.377	.000	1.64	.009007	.367	.021833	.023976
1.35	.367	.027600	.035773	.040337	1.65	.723	.083	.430	.512
1.36	.176	.281	.223	.039680	1.66	.560	.016783	.032	.055-
1.37	.011966	.026384	.031678	.030	1.67	.399	.186	.020638	.022604
1.38	.757	.190	.137	.038386	1.68	.210	.193	.250+	.159
1.39	.550+	.099	.033601	.037719	1.69	.082	.015903	.019968	.021721
1.40	.345-	.025711	.070	.118	1.70	.008927	.617	.490	.290
1.41	.140	.326	.032513	.036493	1.71	.773	.334	.117	.020865-
1.42	.013938	.024914	.022	.035875+	1.72	.621	.055+	.018750-	.446
1.43	.717	.566	.031505+	.264	1.73	.470	.011780	.388	.033
1.44	.537	.190	.039994	.031659	1.74	.322	.508	.030	.018627
1.45	.339	.023818	.487	.061	1.75	.175-	.239	.017678	.227
1.46	.142	.449	.025985+	.031470	1.76	.030	.013974	.331	.018833
1.47	.012817	.084	.189	.032085+	1.77	.007837	.713	.016989	.446
1.48	.754	.022721	.028997	.308	1.78	.745+	.454	.651	.064
1.49	.562	.362	.511	.031736	1.79	.605+	.200	.319	.017689
1.50	.372	.006	.029	.172	1.80	.467	.012949	.015992	.320

TABLE A OF $\pi(h, a)$

h/a	.25	.50	.75	1.00	h/a	.25	.50	.75	1.00
1.80	.007167	.012919	.015992	.017320	2.30	.002625+	.001331	.005079	.005365-
1.81	.331	.701	.669	.016956	2.32	.505-	.186	.004825+	.034
1.82	.197	.457	.352	.599	2.34	.388	.003926	.582	.001714
1.83	.064	.216	.039	.217	2.36	.277	.735-	.127	.527
1.84	.006933	.011978	.014731	.015901	2.38	.169	.551	.350-	.291
1.85	.804	.714	.428	.561	2.40	.066	.375-	.003915-	.065+
1.86	.676	.514	.130	.227	2.42	.001867	.206	.003950+	.003950+
1.87	.550+	.286	.013836	.014898	2.44	.872	.044	.517	.635-
1.88	.426	.062	.517	.575+	2.46	.781	.002890	.332	.419
1.89	.304	.010811	.263	.258	2.48	.693	.712	.155+	.263
1.90	.183	.624	.012983	.013916	2.50	.609	.600	.002987	.086
1.91	.084	.410	.708	.639	2.52	.529	.465+	.826	.002917
1.92	.009917	.199	.437	.338	2.54	.452	.336	.673	.755
1.93	.831	.009591	.171	.042	2.56	.379	.223	.527	.603
1.94	.717	.786	.011909	.012792	2.58	.308	.095+	.388	.455
1.95	.605+	.585-	.652	.467	2.60	.241	.001983	.256	.320
1.96	.485+	.387	.392	.187	2.62	.177	.876	.130	.189
1.97	.386	.192	.150-	.011911	2.64	.115+	.771	.011	.064
1.98	.278	.000	.010905+	.611	2.66	.057	.617	.001897	.001916
1.99	.172	.008811	.665+	.376	2.68	.001	.585-	.789	.834
2.00	.068	.625+	.429	.116	2.70	.000917	.497	.687	.727
2.02	.001865-	.263	.009970	.010611	2.72	.896	.413	.590	.627
2.04	.667	.007912	.577	.124	2.74	.818	.334	.498	.531
2.06	.476	.573	.082	.009656	2.76	.802	.258	.440	.441
2.08	.291	.245+	.008687	.205+	2.78	.758	.186	.327	.355+
2.10	.112	.006929	.291	.008773	2.80	.716	.118	.219	.274
2.12	.003939	.624	.007909	.357	2.82	.676	.054	.175-	.198
2.14	.771	.330	.512	.007958	2.84	.638	.104	.155+	.125+
2.16	.610	.046	.188	.575-	2.86	.602	.934	.038	.057
2.18	.453	.005772	.006819	.207	2.88	.568	.879	.000975-	.000992
2.20	.303	.509	.523	.006855+	2.90	.535+	.827	.916	.931
2.22	.157	.255+	.209	.517	2.92	.504	.777	.859	.874
2.24	.017	.011	.005909	.194	2.94	.475+	.731	.806	.819
2.26	.002881	.001777	.620	.005884	2.96	.448	.686	.756	.768
2.28	.751	.551	.344	.588	2.98	.421	.645-	.709	.720
2.30	.625+	.334	.079	.305-	3.00	.396	.605+	.665-	.674

TABLE B OF $T(h, \bullet)$

$\frac{h}{n}$	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
.00	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
.01	.001591	.001513	.001404	.001201	.965+	.729	.517	.344	.215+	.127	.070	.036	.018
.02	.003183	.003088	.002890	.002492	.001930	.001157	.000333	.000	.431	.253	.140	.073	.035+
.03	.004773	.004626	.004212	.003593	.002696	.00185-	.000932	.000	.645	.379	.210	.109	.053
.04	.006363	.006167	.005615-	.004802	.003898	.00285-	.001932	.375-	.860	.506	.279	.145-	.071
.05	.007951	.007706	.007016	.006000	.004821	.003538	.002579	.717	.001074	.631	.348	.181	.088
.06	.009538	.009244	.008416	.007197	.005782	.004363	.003092	.002059	.588	.757	.417	.216	.105+
.07	.011123	.010780	.009814	.008392	.006741	.005086	.003604	.359	.500+	.881	.495	.252	.123
.08	.012705+	.012314	.011209	.009565-	.007696	.005807	.004115-	.739	.712	.001005+	.555-	.287	.140
.09	.014285+	.013845-	.012603	.010775+	.008693	.006527	.004513-	.003077	.923	.129	.622	.322	.157
.10	.015863	.015373	.013993	.011963	.009605+	.007244	.005131	.413	.002133	.251	.690	.357	.174
.11	.017437	.016898	.015380	.013147	.010555-	.008244	.006138	.748	.341	.373	.757	.392	.190
.12	.019008	.018420	.016764	.014328	.011501	.008670	.006513	.004681	.548	.494	.883	.426	.207
.13	.020575-	.019938	.018144	.015506	.012444	.009379	.006938	.412	.754	.614	.888	.459	.223
.14	.022138	.021452	.019521	.016680	.013384	.010084	.007135-	.740	.958	.733	.953	.492	.239
.15	.023697	.022862	.020893	.017850-	.014319	.010857	.007629	.005067	.003160	.850+	.001017	.525+	.255-
.16	.025251	.024467	.022260	.019015-	.015251	.011485-	.008120	.391	.360	.967	.081	.558	.270
.17	.026800	.025968	.023623	.020176	.016178	.012179	.008608	.712	.559	.002082	.143	.590	.285+
.18	.028344	.027463	.024980	.021331	.017100	.012909	.009092	.006031	.755+	.308	.285-	.621	.300
.19	.029883	.028953	.026333	.022482	.018018	.013555+	.009573	.347	.950-	.308	.266	.652	.315-
.20	.031416	.030437	.027679	.023627	.019050+	.014237	.010050+	.659	.004142	.418	.325+	.628	.329
.21	.032944	.031916	.029020	.024766	.019837	.014637	.010437	.969	.332	.528	.384	.712	.343
.22	.034465+	.033388	.030355-	.025899	.020739	.015585+	.011456	.756	.520	.635+	.442	.741	.357
.23	.035980	.034854	.031683	.027027	.021635-	.016252	.011456	.579	.705+	.741	.499	.769	.371
.24	.037488	.036313	.033005+	.028148	.022525-	.016852	.011456	.879	.888	.846	.555-	.797	.384
.25	.038990	.037766	.034320	.029262	.023408	.017569	.012372	.008175-	.005068	.948	.609	.824	.396
.26	.040484	.039211	.035628	.030370	.024286	.018219	.012746	.008175-	.005068	.948	.609	.824	.396
.27	.041971	.040649	.036989	.031470	.025156	.018419	.012746	.756	.421	.528	.663	.851	.409
.28	.043451	.042080	.038223	.032564	.026021	.019502	.013504	.592	.593	.635+	.716	.877	.421
.29	.044923	.043503	.039509	.033650-	.027027	.020135-	.014147	.322	.762	.741	.817	.927	.444
.30	.046387	.044918	.040787	.034728	.027728	.020135-	.014147	.599	.929	.866	.866	.951	.455-

TABLE B OF T(a, b)

a \ b	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
.30	.046387	.044918	.040787	.034728	.027728	.020761	.014578	.009599	.005929	.003434	.001866	.000951	.000455-
.31	.047803	.046326	.042057	.035799	.028571	.021381	.015003	.009592	.006092	.003592	.001914	.000974	.000476+
.32	.049231	.047725	.043319	.036862	.029407	.022194	.015642	.009942	.006442	.003942	.002006	.001019	.000486+
.33	.050730	.049115	.044573	.037916	.030335	.023001	.016248	.010248	.006648	.004048	.002112	.001059	.000495+
.34	.052261	.050497	.045754	.038863	.031056	.023801	.016848	.010548	.006848	.004148	.002212	.001112	.000505-
.35	.053883	.051871	.046954	.040001	.031868	.024295	.017050	.010548	.006848	.004148	.002212	.001112	.000513
.36	.055597	.053235	.048082	.040673	.032173	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000522
.37	.057401	.054591	.049152	.041382	.032573	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000530
.38	.059297	.055937	.050211	.042058	.032849	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000538
.39	.061283	.057473	.051412	.042868	.033240	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000546
.40	.063359	.059102	.052602	.043573	.033576	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000553
.41	.065527	.060821	.053854	.044382	.033882	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000560
.42	.067789	.062624	.055190	.045258	.034258	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000567
.43	.070147	.064533	.056621	.046258	.034758	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000573
.44	.072591	.066518	.058188	.047382	.035382	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000579
.45	.075129	.068598	.059818	.048518	.036018	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000585
.46	.077761	.070768	.061568	.050068	.037268	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000590
.47	.080489	.073066	.063466	.051466	.038266	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000595
.48	.083317	.075466	.065466	.052966	.039466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000600
.49	.086245	.077966	.067466	.054466	.040466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000605
.50	.089273	.080466	.069466	.055966	.041466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000610
.51	.092401	.083166	.071466	.057466	.042466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000615
.52	.095629	.085966	.073766	.059466	.043966	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000620
.53	.098957	.088766	.076166	.061466	.045466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000625
.54	.102385	.091666	.078666	.063466	.047466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000630
.55	.105913	.094666	.081166	.065466	.049466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000635
.56	.109541	.097966	.083966	.068466	.052466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000640
.57	.113269	.101466	.086966	.071466	.054966	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000645
.58	.117097	.104666	.090166	.074466	.058466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000650
.59	.121025	.108466	.093466	.078466	.062466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000655
.60	.125053	.112466	.096966	.082466	.066466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000660
.61	.129181	.116466	.100466	.086466	.070466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000665
.62	.133409	.120466	.104466	.090466	.074466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000670
.63	.137737	.124466	.108466	.094466	.078466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000675
.64	.142165	.128466	.112466	.098466	.082466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000680
.65	.146693	.132466	.116466	.102466	.086466	.024173	.017050	.010548	.006848	.004148	.002212	.001112	.000685

TABLE B OF $T(h, a)$

λ	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
.65	.091733	.088557	.079681	.064839	.052291	.038177	.026028	.016935	.009895	.005516	.003884	.002415	.000651
.66	.092807	.089622	.080681	.065839	.053291	.039177	.027028	.017935	.010895	.006516	.004884	.003415	.001651
.67	.093950	.090765	.081824	.066982	.054434	.040320	.028171	.019078	.012038	.007659	.006027	.004558	.002794
.68	.095093	.091908	.082967	.068125	.055577	.041463	.029314	.020221	.013181	.008802	.007170	.005701	.003937
.69	.096236	.093051	.084110	.069268	.056720	.042606	.030457	.021364	.014324	.009945	.008313	.006844	.005080
.70	.097379	.094194	.085253	.070411	.057863	.043749	.031600	.022507	.015467	.011088	.009456	.007987	.006223
.71	.098522	.095337	.086396	.071554	.059006	.044892	.032743	.023650	.016610	.012231	.010599	.009130	.007366
.72	.099665	.096480	.087539	.072697	.060149	.046035	.033886	.024793	.017753	.013374	.011742	.010273	.008509
.73	.100808	.097623	.088682	.073840	.061292	.047178	.035029	.025936	.018896	.014517	.012885	.011416	.009652
.74	.101951	.098766	.089825	.074983	.062435	.048321	.036172	.027079	.020039	.015660	.014028	.012559	.010795
.75	.103094	.099909	.090968	.076126	.063578	.049464	.037315	.028222	.021182	.016803	.015171	.013702	.011938
.76	.104237	.101052	.092111	.077269	.064721	.050607	.038458	.029365	.022325	.017946	.016314	.014845	.013081
.77	.105380	.102195	.093254	.078412	.065864	.051750	.039601	.030508	.023468	.019089	.017457	.015988	.014224
.78	.106523	.103338	.094397	.079555	.067007	.052893	.040744	.031651	.024611	.020232	.018600	.017131	.015367
.79	.107666	.104481	.095540	.080698	.068150	.054036	.041887	.032794	.025754	.021375	.019743	.018274	.016510
.80	.108809	.105624	.096683	.081841	.069293	.055179	.043030	.033937	.026897	.022518	.020886	.019417	.017653
.81	.109952	.106767	.097826	.082984	.070436	.056322	.044173	.035080	.028040	.023661	.022029	.020560	.018796
.82	.111095	.107910	.098969	.084127	.071579	.057465	.045316	.036223	.029183	.024804	.023172	.021703	.019939
.83	.112238	.109053	.100112	.085270	.072722	.058608	.046459	.037366	.030326	.025947	.024315	.022846	.021082
.84	.113381	.110196	.101255	.086413	.073865	.059751	.047602	.038509	.031469	.027090	.025458	.023989	.022225
.85	.114524	.111339	.102398	.087556	.075008	.060894	.048745	.039652	.032612	.028233	.026601	.025132	.023368
.86	.115667	.112482	.103541	.088699	.076151	.062037	.050888	.041795	.034755	.030376	.028744	.027275	.025511
.87	.116810	.113625	.104684	.089857	.077309	.063195	.052046	.042953	.035913	.031534	.029902	.028433	.026669
.88	.117953	.114768	.105827	.091025	.078477	.064363	.053214	.044121	.037081	.032702	.031070	.029601	.027837
.89	.119096	.115911	.106970	.092168	.079620	.065506	.054357	.045264	.038224	.033845	.032213	.030744	.028980
.90	.120239	.117054	.108113	.093311	.080763	.066649	.055500	.046407	.039367	.034988	.033356	.031887	.030123
.91	.121382	.118197	.109256	.094454	.081906	.067792	.056643	.047550	.040510	.036131	.034500	.033031	.031267
.92	.122525	.119340	.110399	.095597	.083049	.068935	.057786	.048693	.041653	.037274	.035643	.034174	.032410
.93	.123668	.120483	.111542	.096740	.084192	.070078	.058929	.049836	.042796	.038417	.036786	.035317	.033553
.94	.124811	.121626	.112685	.097883	.085335	.071221	.060072	.050979	.043939	.039560	.037929	.036460	.034696
.95	.125954	.122769	.113828	.099026	.086478	.072364	.061215	.052122	.045082	.040703	.039072	.037603	.035839
.96	.127097	.123912	.114971	.100169	.087621	.073507	.062358	.053265	.046225	.041846	.040215	.038746	.036982
.97	.128240	.125055	.116114	.101312	.088764	.074650	.063501	.054408	.047368	.042989	.041358	.039889	.038125
.98	.129383	.126198	.117257	.102455	.090907	.076793	.065644	.056551	.049511	.045132	.043501	.042032	.040268
.99	.130526	.127341	.118400	.103598	.092050	.077936	.066787	.057694	.050654	.046275	.044644	.043175	.041411
1.00	.131669	.128484	.119543	.104741	.093193	.079079	.067930	.058837	.051797	.047418	.045787	.044318	.042554
∞	.132812	.129627	.120686	.105884	.094336	.080222	.069073	.060980	.053940	.049561	.047930	.046461	.044697

TABLE C OF T(h, a)

[illegible]

may be used throughout Table C. Tables A and B were designed for interpolation as follows: To interpolate for a value $T(h_2, a_2)$, say, a_1 and a_3 should be picked closest to a_2 from Table A so that $a_1 \leq a_2 < a_3$, and h_1 and h_3 should be picked closest to h_2 from Table B so that $h_1 \leq h_2 < h_3$. Then the interpolated value of $T(h_2, a_2)$ is obtained from

$$T(h_2, a_2) = \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} T(h_i, a_j),$$

where the weights w_{ij} are given by

$$w_{ij} = \begin{pmatrix} -(1-b)(1-c) & 1-c & -b(1-c) \\ (1-b) & 0 & b \\ -(1-b)c & c & -bc \end{pmatrix},$$

where

$$b = \frac{a_2 - a_1}{a_3 - a_1} \quad \text{and} \quad c = \frac{h_2 - h_1}{h_3 - h_1}.$$

The weights were obtained by considering the result of ordinary linear interpolation where nearby values of the function are subtracted before interpolating, say, $T(h_2, a_1) - T(h_1, a_1)$ and $T(h_2, a_3) - T(h_1, a_3)$. These numbers are interpolated with respect to a to obtain $T(h_2, a_2) - T(h_1, a_2)$, and then $T(h_1, a_2)$ is added. This process may also be followed with (h_3, a_1) , (h_3, a_3) , and (h_3, a_2) . If the two estimates of $T(h_2, a_2)$ are then combined as in linear interpolation with respect to h , i.e., $(1-c)$ times the first estimate plus c times the second, the above weights w_{ij} follow. The interpolation on the differences could also have been first with respect to h to obtain the two estimates and then with respect to a between these two. The same weights w_{ij} are obtained by doing this.

This method of interpolation has resulted in approximately a 90 per cent reduction over the size of a table needed for linear interpolation. Quadratic interpolation using Bessel's formula would give comparable results to the new method with approximately an additional 80 per cent reduction in the number of entries, but the additional work involved more than outweighs that reduction in the number of entries, even though the table is used only a few times. The procedure given here may be termed a compromise between linear and quadratic interpolation.

EXAMPLE. Find $T(.15, .625)$. From the tables, the following entries are extracted:

h	a		
	.50	.625	.75
0	.073792	.088903	.102416
.15	.072902		.101082
.25	.071347	.085848	.098755

The weights to be applied are

h	a		
	.50	.625	.75
0	-.2	.4	-.2
.15	.5		.5
.25	-.3	.6	-.3

The result is $T(.15, .625) = .0877898$. Calculation of this number from the series gives .0877919. The result of the interpolation therefore provides a difference of two in the sixth place. Further calculations show that this difference could be reduced to five in the seventh place if the linear interpolations for $T(0, .625)$ and $T(.25, .625)$ were eliminated and the exact values for these points were used. The value for $T(0, .625)$ was rounded up during the linear interpolation with respect to a in Table B, since second differences in the a direction for all h are negative. A similar working rule for rounding when interpolating in the h direction is to round up the interpolated value when $0 < h < .9$ and to round down for $h \geq .9$ in Table A. The value obtained from the above interpolation scheme should be rounded up for $0 < h < 1.50$ and rounded down for $h \geq 1.50$, for all values of $a > 0$.

Empirical examination of the errors in interpolation by this scheme shows that the maximum error that would occur anywhere in Tables A and B is seven in the sixth decimal place, and that this could be reduced to six in the sixth decimal place if the linear interpolations in Tables A and B were eliminated and the exact values were used. Linear interpolation in Table C gives errors less than four in the sixth decimal place. Table D gives the maximum error in the sixth decimal place, which will be committed when using the above

TABLE D

h	a			
	0.00-0.25	0.25-0.50	0.50-0.75	0.75-1.00
0.00-0.25	+0.9	+2.1	+2.8	+3.3
0.25-0.50	+1.3	+3.1	+4.4	+5.3
0.50-0.75	+1.7	+4.4	+6.2	+7.1
0.75-1.00	+2.0	+4.9	+6.5	+6.4
1.00-1.25	+1.8	+4.2	+4.6	+3.1
1.25-1.50	+1.3	+2.5	+1.5	-1.3
1.50-1.75	+0.5	+0.5	-1.7	-3.9
1.75-2.00	-0.3	-1.6	-3.6	-4.7
2.00-2.25	-0.8	-2.6	-4.0	-3.9
2.25-2.50	-1.0	-2.8	-3.4	-2.6
2.50-2.75	-1.0	-2.4	-2.4	-1.4
2.75-3.00	-0.8	-1.7	-1.4	-0.7

interpolation scheme over the ranges of h and a indicated. The sign preceding the entry is the sign of the exact value of $T(h, a)$ minus the interpolated value for that difference which is the largest in absolute value. These are empirical results obtained on the digital computer by using the interpolation process and the exact value for fifteen points systematically spaced in each block. A number in Tables A, B, and C whose last nonzero digit is five is followed by a plus or minus sign, respectively, to indicate that the number should be rounded up or down when dropping the digit with the five. Any entry having the first three digits the same as those of the entry directly above it has had these digits dropped.

REFERENCES

- [1] KARL PEARSON, *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, 1931.
- [2] *Tables of the Bivariate Normal Distribution*, University of California, Statistical Laboratory, Berkeley. (These tables are unpublished; however, a description of these tables is given in *Mathematics Tables and Other Aids to Computation*, Vol. II, 1948, pp. 372-373.)
- [3] D. TEICHHROEW, "Numerical Analysis Research unpublished statistical tables," *J. Amer. Stat. Assn.*, Vol. 50, 1955, pp. 550-556.
- [4] W. F. SHEPPARD, "II. On the calculation of the double-integral expressing normal correlation," *Trans. Cambridge Phil. Soc.*, Vol. 19, 1900, pp. 23-66.
- [5] C. NICHOLSON, "The probability integral for two variables," *Biometrika*, Vol. 33, 1943, pp. 59-72.
- [6] R. C. T. SMITH, "Conduction of heat in the semi-infinite solid, with a short table of an important integral," *Australian Journal of Physics*, Vol. 6, 1953, pp. 127-130.
- [7] C. HASTINGS AND M. PIEDER, *Miscellaneous Probability Tables*, Applied Mathematics Panel, National Defense Research Committee, Note No. 14, New York, 1944.
- [8] A. R. EAGLE, "A method for handling errors in testing and measuring," *Industrial Quality Control*, Vol. 10, 1954, pp. 10-15.
- [9] A. H. BOWKER AND H. P. GOODE, *Sampling Inspection by Variables*, McGraw-Hill Book Company, 1952, pp. 127-132.
- [10] A. R. KAMAT, "Incomplete and absolute moments of the multivariate normal distribution with some applications," *Biometrika*, Vol. 40, 1953, pp. 20-34.
- [11] J. H. CADWELL, "The bivariate normal integral," *Biometrika*, Vol. 38, 1951, pp. 475-481.
- [12] H. CRAMER, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 290.
- [13] F. N. DAVID, "A note on the evaluation of the multivariate normal integral," *Biometrika*, Vol. 40, 1953, pp. 458-459.
- [14] R. L. PLACKETT, "A reduction formula for normal multivariate integrals," *Biometrika*, Vol. 41, 1954, pp. 351-360.
- [15] H. A. BENDER, "Bivariate distribution," *Bull. Amer. Math. Soc.*, Vol. 61, abstract No. 813, pp. 561-562.

ASYMPTOTIC FORMULAE FOR THE DISTRIBUTION OF HOTELLING'S GENERALIZED T_0^2 STATISTIC¹

BY KOICHI ITO²

University of North Carolina

1. Summary. In this paper the asymptotic expansion of a percentage point of Hotelling's generalized T_0^2 distribution is derived in terms of the corresponding percentage point of a χ^2 distribution. Our result generalizes Hotelling's and Frankel's asymptotic expansion for the generalized Student T [3], [4]. The technique used in this paper for obtaining the asymptotic expansion of T_0^2 is an extension of the previous methods of Welch [8] and of James [5], [6], who used them to solve the distribution problem of various statistics in connection with the Behrens-Fisher problem. An asymptotic formula for the cumulative distribution function (c.d.f.) of T_0^2 is also given together with an upper bound for the error committed when all but the first few terms are omitted in the series. This formula is a sort of multivariate analogue of Hartley's formula of "Studentization" [2].

2. Introduction. In the multivariate analysis of variance we use the following canonical probability law:

$$(2.1) \quad P(X_0, X_1) = \text{const.} \exp \left[-\frac{1}{2} \text{tr} \Lambda(X_1 - \xi)(X_1' - \xi') - \frac{1}{2} \text{tr} \Lambda X_0 X_0' \right] dX_0 dX_1,$$

where X_1 and X_0 are $p \times m$ and $p \times m$ matrices, respectively, and $(1/m)X_1 X_1' = S_1$ is the sample "between" dispersion matrix and $(1/n)X_0 X_0' = S_0$ is the sample "within" dispersion matrix, the prime denoting the transpose of a matrix. ξ is a $p \times m$ matrix, $(1/m)\xi\xi'$ being the population "between" dispersion matrix, and Λ is a $p \times p$ symmetric positive definite matrix. It is assumed that m may be $\geq p$ or $< p$, but $n \geq p$. To test the null hypothesis $H_0: \xi = 0$, Hotelling [3] proposed a test based on the statistic:

$$(2.2) \quad T_0^2 = m \text{tr} S_1 S_0^{-1}$$

and derived the exact distribution of this statistic when $p = 2$ and $\xi = 0$. For general values of p the exact distribution of T_0^2 is not available at present, even in the null case $\xi = 0$.

3. Derivation of asymptotic formula of T_0^2 . For general values of p it is known that the statistic

$$(3.1) \quad \chi^2 = m \text{tr} S_1 \Lambda$$

Received July 6, 1955.

¹ Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

² Now at Nanzan University, Nagoya, Japan.

has a χ^2 distribution with mp degrees of freedom. That is to say, we have

$$(3.2) \quad \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\} = G_\rho(\theta),$$

where 2θ denotes the tabled value of χ^2 for a particular level of significance, $\rho = mp/2$, and

$$G_\rho(\theta) = [\Gamma(\rho)]^{-1} \int_0^\theta t^{\rho-1} e^{-t} dt.$$

Hence, if Λ is known, the statistic χ^2 given by (3.1) may be used to test H_0 exactly, and if Λ is unknown but if S_0 is based on a large number of degrees of freedom, i.e., if n is large, we may use as an approximation the result

$$(3.4) \quad \Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2\theta\} = G_\rho(\theta).$$

This suggests that in the general case we try to find a function $h(S_0)$ of the elements of S_0 such that

$$(3.5) \quad \Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0)\} = G_\rho(\theta).$$

When n is large, $2h(S_0)$ will approach $2\theta \equiv \chi^2$, and we now expect to write $2h(S_0)$ as a series with χ^2 as its first term and successive terms of decreasing order of magnitude.

Now

$$(3.6) \quad \Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0)\} = \int_R \Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0) \mid S_0\} \Pr \{dS_0\},$$

where the first expression on the right denotes the conditional probability of the relation indicated for fixed values of the elements of S_0 , and the second denotes the probability element of S_0 , which has a Wishart distribution with n degrees of freedom, and the domain of integration R is over all possible values of the elements of S_0 . Now we may expand $\Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0) \mid S_0\}$ about an origin $(\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp}, \sigma_{12}, \dots, \sigma_{p-1,p})$ in a Taylor series, where

$$\Lambda^{-1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}.$$

Thus,

$$\begin{aligned} & \Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0) \mid S_0\} \\ (3.7) \quad &= \left\{ \exp \left[\sum_{i,j=1}^p (s_{0ij} - \sigma_{ij}) \frac{\partial}{\partial \sigma_{ij}} \right] \right\} \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2h(\Lambda^{-1})\} \\ &= \{\exp [\operatorname{tr} (S_0 - \Lambda^{-1}) \partial]\} \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2h(\Lambda^{-1})\}, \end{aligned}$$

where s_{0ij} is the i th row, j th column element of S_0 , and ∂ denotes the matrix of derivative operators:

$$(3.8) \quad \partial = \begin{bmatrix} \frac{\partial}{\partial \sigma_{11}} & \frac{1}{2} \frac{\partial}{\partial \sigma_{12}} & \cdots & \frac{1}{2} \frac{\partial}{\partial \sigma_{1p}} \\ \frac{1}{2} \frac{\partial}{\partial \sigma_{21}} & \frac{\partial}{\partial \sigma_{22}} & \cdots & \frac{1}{2} \frac{\partial}{\partial \sigma_{2p}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} \frac{\partial}{\partial \sigma_{p1}} & \frac{1}{2} \frac{\partial}{\partial \sigma_{p2}} & \cdots & \frac{\partial}{\partial \sigma_{pp}} \end{bmatrix},$$

its typical element being $\partial_{ij} = \frac{1}{2}(1 + \delta_{ij})(\partial/\partial \sigma_{ij})$, where δ_{ij} is the Kronecker delta. Whether uniformly convergent or not, the right-hand side of (3.7) is an asymptotic representation of $\Pr \{m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0) \mid S_0\}$, for sufficiently large values of n . Hence, substitution of (3.7) into (3.6) and term by term integration, which may be done legitimately, yields:

$$(3.9) \quad \begin{aligned} G_p(\theta) &= \int_R \exp [\operatorname{tr} (S_0 - \Lambda^{-1})\partial] \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2h(\Lambda^{-1})\} \Pr \{dS_0\} \\ &= \Theta \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2h(\Lambda^{-1})\}, \end{aligned}$$

where

$$\Theta = \int_R \exp [\operatorname{tr} (S_0 - \Lambda^{-1})\partial] \Pr \{dS_0\}.$$

Since S_0 has a Wishart distribution with n degrees of freedom, we have

$$\begin{aligned} \Theta &= \exp [-\operatorname{tr} \Lambda^{-1} \partial] \cdot \text{const.} \cdot |\Lambda|^{n/2} \int_R |S_0|^{(n-p-1)/2} \\ &\quad \cdot \exp \left[\operatorname{tr} \left(S_0 \partial - \frac{n}{2} \Lambda S_0 \right) \right] dS_0 \\ &= \exp [-\operatorname{tr} \Lambda^{-1} \partial] \cdot \text{const.} \cdot |\Lambda|^{n/2} \int_R |S_0|^{(n-p-1)/2} \\ &\quad \cdot \exp \left[-\frac{n}{2} \operatorname{tr} \left(\Lambda - \frac{2}{n} \partial \right) S_0 \right] dS_0 \\ &= \exp [-\operatorname{tr} \Lambda^{-1} \partial] \cdot |\Lambda|^{n/2} \left| \Lambda - \frac{2}{n} \partial \right|^{-n/2} \\ &= \exp [-\operatorname{tr} \Lambda^{-1} \partial] \cdot \left| I - \frac{2}{n} \Lambda^{-1} \partial \right|^{-n/2}, \end{aligned}$$

where I is the $p \times p$ identity matrix. Now using [5],

$$(3.10) \quad -\log |I - Y| = \operatorname{tr} Y + \frac{1}{2} \operatorname{tr} Y^2 + \frac{1}{3} \operatorname{tr} Y^3 + \cdots,$$

we obtain

$$\begin{aligned}
 \Theta &= \exp \left[-\text{tr } \Lambda^{-1} \partial - \frac{n}{2} \log \left| I - \frac{2}{n} \Lambda^{-1} \partial \right| \right] \\
 &= \exp \left[-\text{tr } \Lambda^{-1} \partial + \frac{n}{2} \left\{ \text{tr} \left(\frac{2}{n} \Lambda^{-1} \partial \right) + \frac{1}{2} \text{tr} \left(\frac{2}{n} \Lambda^{-1} \partial \right)^2 \right. \right. \\
 (3.11) \quad &\quad \left. \left. + \frac{1}{3} \text{tr} \left(\frac{2}{n} \Lambda^{-1} \partial \right)^3 + \cdots \right\} \right] \\
 &= \exp \left[\frac{1}{n} \text{tr} (\Lambda^{-1} \partial)^2 + \frac{4}{3n^2} \text{tr} (\Lambda^{-1} \partial)^3 + \cdots \right] \\
 &= 1 + \frac{1}{n} \text{tr} (\Lambda^{-1} \partial)^2 + \frac{1}{n^2} \left\{ \frac{4}{3} \text{tr} (\Lambda^{-1} \partial)^3 + \frac{1}{2} (\text{tr} (\Lambda^{-1} \partial)^2)^2 \right\} + O(n^{-3}).
 \end{aligned}$$

It is to be noted here that in (3.11) the operator ∂ does not act on Λ^{-1} present in Θ itself, and it is more useful for our purpose to write (3.11) in suffix form:

$$\begin{aligned}
 \Theta &= 1 + \frac{1}{n} \sum \sigma_{rs} \sigma_{tu} \partial_{st} \partial_{ur} \\
 (3.12) \quad &+ \frac{1}{n^2} \left\{ \frac{4}{3} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \partial_{st} \partial_{uv} \partial_{wr} + \frac{1}{2} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} \partial_{st} \partial_{ur} \partial_{wx} \partial_{yv} \right\} \\
 &+ O(n^{-3}),
 \end{aligned}$$

where \sum denotes the summation over all suffixes r, s, \dots , each of which ranges from 1 to p .

Now we represent $h(S_0)$ as

$$(3.13) \quad h(S_0) = \theta + h_1(S_0) + h_2(S_0) + \cdots,$$

$h_s(S_0)$ being of order n^{-s} ; i.e., we write $h(S_0)$ as an asymptotic series such that

$$|n^s \{h(S_0) - \theta - h_1(S_0) - \cdots - h_s(S_0)\}|$$

is made arbitrarily small for sufficiently large values of n . Then (3.13) may be substituted into $\text{Pr} \{m \text{tr } S_1 \Lambda \leq 2h(\Lambda^{-1})\}$, and by Taylor's expansion we have

$$\begin{aligned}
 \text{Pr} \{m \text{tr } S_1 \Lambda \leq 2h(\Lambda^{-1})\} \\
 &= \exp \{[h_1(\Lambda^{-1}) + h_2(\Lambda^{-1}) + \cdots] D\} \text{Pr} \{m \text{tr } S_1 \Lambda \leq 2\theta\} \\
 (3.14) \quad &= [1 + \{h_1(\Lambda^{-1}) + h_2(\Lambda^{-1}) + \cdots\} D \\
 &\quad + \frac{1}{2} \{h_1(\Lambda^{-1}) + h_2(\Lambda^{-1}) + \cdots\}^2 D^2 + \cdots] \\
 &\quad \times \text{Pr} \{m \text{tr } S_1 \Lambda \leq 2\theta\},
 \end{aligned}$$

where $D = \partial/\partial\theta$. By substituting (3.12) and (3.14) into (3.9), we obtain

$$\begin{aligned}
 G_p(\theta) = & \left[1 + \frac{1}{n} \sum \sigma_{rs} \sigma_{tu} \partial_{st} \partial_{ur} \right. \\
 & + \frac{1}{n^2} \left\{ \frac{4}{3} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \partial_{st} \partial_{uv} \partial_{wr} \right. \\
 (3.15) \quad & \left. + \frac{4}{3} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} \partial_{st} \partial_{ur} \partial_{wx} \partial_{yv} \right\} + O(n^{-3}) \Big] \\
 & \times [1 + h_1(\Lambda^{-1}) D + \{h_2(\Lambda^{-1}) D + \frac{1}{2} h_1^2(\Lambda^{-1}) D^2\} + O(n^{-3})] \\
 & \times \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\}.
 \end{aligned}$$

By equating terms of successive order in (3.15), we obtain

$$\begin{aligned}
 (3.16) \quad & \left\{ h_1(\Lambda^{-1}) D + \frac{1}{n} \sum \sigma_{rs} \sigma_{tu} \partial_{st} \partial_{ur} \right\} \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\} = 0, \\
 (3.17) \quad & \left[h_2(\Lambda^{-1}) D + \frac{1}{2} h_1^2(\Lambda^{-1}) D^2 \right. \\
 & + \frac{1}{n} \sum \sigma_{rs} \sigma_{tu} \{h_1^{(st,ur)}(\Lambda^{-1}) D + 2h_1^{(st)}(\Lambda^{-1}) \partial_{ur} D + h_1(\Lambda^{-1}) \partial_{st} \partial_{ur} D\} \\
 & + \frac{4}{3n^2} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \partial_{st} \partial_{uv} \partial_{wr} + \frac{1}{2n^2} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} \partial_{st} \partial_{ur} \partial_{wx} \partial_{yv} \Big] \\
 & \times \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\} = 0,
 \end{aligned}$$

and so on, where $h_1^{(st)}(\Lambda^{-1}) = \partial_{st} h_1(\Lambda^{-1})$ and $h_1^{(st,ur)}(\Lambda^{-1}) = \partial_{ur} \partial_{st} h_1(\Lambda^{-1})$.

It now remains to carry out the operations ∂ and D indicated in (3.16) and (3.17) in order to obtain $h_1(\Lambda^{-1})$, $h_2(\Lambda^{-1})$ and hence $h_1(S_0)$, $h_2(S_0)$. These operators will operate on $\Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\}$, which is a $p \times m$ -fold integral, and the operations may be thought of as differentiations, with respect to the boundary only, of the integral of the probability density function of the X_1 throughout a region in the space of X_1 . The method used to evaluate $\partial_{st} \partial_{ur} \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\}$, $\partial_{st} \partial_{ur} \partial_{wr} \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\}$, \dots , is to change the boundary slightly, expand the integral in powers of the quantities specifying this change, and obtain the derivatives by comparison with Taylor's expansion. We consider

$$(3.18) \quad J = \Pr \{m \operatorname{tr} S_1 (\Lambda^{-1} + \epsilon)^{-1} \leq 2\theta\},$$

where ϵ is a $p \times p$ symmetric matrix. Then by Taylor expansion we have

$$\begin{aligned}
 (3.19) \quad J = & \left\{ 1 + \sum \epsilon_{rs} \partial_{rs} + \frac{1}{2!} \sum \epsilon_{rs} \epsilon_{tu} \partial_{rs} \partial_{tu} + \frac{1}{3!} \sum \epsilon_{rs} \epsilon_{tu} \epsilon_{vw} \partial_{rs} \partial_{tu} \partial_{vw} \right. \\
 & \left. + \frac{1}{4!} \sum \epsilon_{rs} \epsilon_{tu} \epsilon_{vw} \epsilon_{xy} \partial_{rs} \partial_{tu} \partial_{vw} \partial_{xy} + \dots \right\} \Pr \{m \operatorname{tr} S_1 \Lambda \leq 2\theta\}.
 \end{aligned}$$

On the other hand, J is, by definition, written as

$$(3.20) \quad J = \frac{|\Lambda|^{m/2}}{(2\pi)^{pm/2}} \int_{R'} \exp \left[-\frac{1}{2} \operatorname{tr} \Lambda X_1 X_1' \right] dX_1,$$

where $X_1 X_1' = m S_1$, and domain of integration R' ranges over all possible values of the elements of X_1 such that $m \operatorname{tr} S_1 (\Lambda^{-1} + \epsilon)^{-1} \leq 2\theta$. It is now easy to show that integration of (3.20) yields

$$(3.21) \quad J = \left(\frac{|I - D_s E|}{|I - D_s|} \right)^{-m/2} G_s(\theta),$$

where D_s is a diagonal matrix which satisfies

$$(3.22) \quad \begin{aligned} X_1(p \times m) &= \Gamma(p \times p) Z(p \times m), \\ \frac{1}{2} \Gamma' (\Lambda^{-1} + \epsilon)^{-1} \Gamma &= I(p), \end{aligned}$$

and

$$\frac{1}{2} \Gamma' \Lambda \Gamma = I(p) - D_s,$$

Γ being a nonsingular matrix, and E is an operator such that

$$E G_s(\theta) = G_{s+1}(\theta).$$

Now, letting $\Delta = E - 1$ and using (3.22), we have

$$\begin{aligned} \frac{|I - D_s E|}{|I - D_s|} &= \frac{|I - D_s - D_s \Delta|}{|I - D_s|} \\ &= \frac{|\frac{1}{2} \Gamma' \Lambda \Gamma - \{\frac{1}{2} \Gamma' (\Lambda^{-1} + \epsilon)^{-1} \Gamma - \frac{1}{2} \Gamma' \Lambda \Gamma\} \Delta|}{|\frac{1}{2} \Gamma' \Lambda \Gamma|} \\ &= \frac{|\Lambda - \{(\Lambda^{-1} + \epsilon)^{-1} - \Lambda\} \Delta|}{|\Lambda|} = |I - \{\Lambda^{-1} (\Lambda^{-1} + \epsilon)^{-1} - I\} \Delta| \\ &= |I - X \Delta|, \end{aligned}$$

where $X = \Lambda^{-1} (\Lambda^{-1} + \epsilon)^{-1} - I$. Hence, (3.21) becomes

$$(3.23) \quad J = |I - X \Delta|^{-m/2} G_s(\theta).$$

Now, using (3.10) again, we rewrite (3.23) as

$$\begin{aligned} J &= \exp \left\{ -\frac{m}{2} \log |I - X \Delta| \right\} G_s(\theta) \\ &= \exp \left\{ \frac{m}{2} \operatorname{tr} X \Delta + \frac{m}{4} \operatorname{tr} X^2 \Delta^2 + \frac{m}{6} \operatorname{tr} X^3 \Delta^3 + \frac{m}{8} \operatorname{tr} X^4 \Delta^4 + \cdots \right\} G_s(\theta) \\ &= \left[1 + \frac{m}{2} \operatorname{tr} X \Delta + \left\{ \frac{m}{4} \operatorname{tr} X^2 + \frac{m^2}{8} (\operatorname{tr} X)^2 \right\} \Delta^2 \right. \end{aligned}$$

$$\begin{aligned}
 (3.24) \quad & + \left\{ \frac{m}{6} \operatorname{tr} X^3 + \frac{m^2}{8} (\operatorname{tr} X)(\operatorname{tr} X^2) + \frac{m^3}{48} (\operatorname{tr} X)^3 \right\} \Delta^3 \\
 & + \left\{ \frac{m}{8} \operatorname{tr} X^4 + \frac{m^2}{12} (\operatorname{tr} X)(\operatorname{tr} X^3) + \frac{m^2}{32} (\operatorname{tr} X^2)^2 \right. \\
 & \left. + \frac{m^3}{32} (\operatorname{tr} X)^2(\operatorname{tr} X^2) + \frac{m^4}{384} (\operatorname{tr} X)^4 \right\} \Delta^4 + \dots \Big] G_p(\theta);
 \end{aligned}$$

X can be represented as

$$\begin{aligned}
 (3.25) \quad X &= \Lambda^{-1}(\Lambda^{-1} + \epsilon)^{-1} - I = \Lambda^{-1}(\Lambda^{-1} + \sum \epsilon_{rs} \Lambda_{rs}^{-1})^{-1} \\
 &- I = (I + \sum \epsilon_{rs} \Lambda_{rs}^{-1})^{-1} - I \\
 &= - \sum \epsilon_{rs} (\Lambda_{rs}^{-1} \Lambda) + \sum \epsilon_{rs} \epsilon_{tu} (\Lambda_{rs}^{-1} \Lambda) (\Lambda_{tu}^{-1} \Lambda) \\
 &- \sum \epsilon_{rs} \epsilon_{tu} \epsilon_{vw} (\Lambda_{rs}^{-1} \Lambda) (\Lambda_{tu}^{-1} \Lambda) (\Lambda_{vw}^{-1} \Lambda) \\
 &+ \sum \epsilon_{rs} \epsilon_{tu} \epsilon_{vw} \epsilon_{xy} (\Lambda_{rs}^{-1} \Lambda) (\Lambda_{tu}^{-1} \Lambda) (\Lambda_{vw}^{-1} \Lambda) (\Lambda_{xy}^{-1} \Lambda) - \dots,
 \end{aligned}$$

where Λ_{rs}^{-1} is a $p \times p$ matrix obtained by operating ∂_{rs} on Λ , i.e., Λ_{rs}^{-1} has its i th row, j th column element, $\frac{1}{2}(\delta_{ri}\delta_{sj} + \delta_{si}\delta_{rj})$. Writing

$$\begin{aligned}
 \operatorname{tr} (\Lambda_{rs}^{-1} \Lambda) &= (rs), \\
 \operatorname{tr} (\Lambda_{rs}^{-1} \Lambda) (\Lambda_{tu}^{-1} \Lambda) &= (rs | tu), \\
 \operatorname{tr} (\Lambda_{rs}^{-1} \Lambda) (\Lambda_{tu}^{-1} \Lambda) (\Lambda_{vw}^{-1} \Lambda) &= (rs | tu | vw), \\
 \operatorname{tr} (\Lambda_{rs}^{-1} \Lambda) (\Lambda_{tu}^{-1} \Lambda) (\Lambda_{vw}^{-1} \Lambda) (\Lambda_{xy}^{-1} \Lambda) &= (rs | tu | vw | xy),
 \end{aligned}$$

and substituting (3.25) into (3.24), we obtain

$$\begin{aligned}
 (3.26) \quad J &= \left[1 + \sum \epsilon_{rs} \left\{ -\frac{m}{2} (rs) \Delta \right\} + \frac{1}{2!} \sum \epsilon_{rs} \epsilon_{tu} \left\{ (rs | tu) \left(m\Delta + \frac{m}{2} \Delta^2 \right) \right. \right. \\
 &+ \frac{m^2}{4} (rs)(tu) \Delta^2 \Big\} + \frac{1}{3!} \sum \epsilon_{rs} \epsilon_{tu} \epsilon_{vw} \left\{ (rs | tu | vw) (-3m\Delta - 3m\Delta^2 - m\Delta^3) \right. \\
 &+ (rs)(tu | vw) \left(-\frac{3}{2} m^2 \Delta^2 - \frac{3}{4} m^2 \Delta^3 \right) - \frac{m^3}{8} (rs)(tu)(vw) \Delta^3 \Big\} \\
 &+ \frac{1}{4!} \sum \epsilon_{rs} \epsilon_{tu} \epsilon_{vw} \epsilon_{xy} \left\{ (rs | tu | vw | xy) (12m\Delta + 18m\Delta^2 + 12m\Delta^3 + 3m\Delta^4) \right. \\
 &+ (rs)(tu | vw | xy) (6m^2 \Delta^2 + 6m^2 \Delta^3 + 2m^2 \Delta^4) \\
 &+ (rs | tu)(vw | xy) (3m^2 \Delta^2 + 3m^2 \Delta^3 + \frac{3}{4} m^2 \Delta^4) \\
 &+ (rs)(tu)(vw | xy) (\frac{3}{2} m^3 \Delta^3 + \frac{3}{4} m^3 \Delta^4) \\
 &\left. \left. + (rs)(tu)(vw)(xy) \frac{m^4}{16} \Delta^4 \right\} + \dots \right] G_p(\theta).
 \end{aligned}$$

Then term by term comparison between two expansions for J , (3.19) and (3.26), gives $\partial_{rs} \Pr \{m \text{ tr } S_1 \Lambda \leq 2\theta\}$, $\partial_{rs} \partial_{tu} \Pr \{m \text{ tr } S_1 \Lambda \leq 2\theta\}$, etc., but in doing so we must take such a care that, for example,

$$\sum a_{ijk} \epsilon_i \epsilon_j \epsilon_k = \sum b_{ijk} \epsilon_i \epsilon_j \epsilon_k$$

implies $a_{ijk} = b_{ijk}$ if both a_{ijk} and b_{ijk} are completely symmetrical in their suffices. With this in mind and using the relation

$$\Delta G_p(\theta) = -E g_p(\theta),$$

where $g_p(\theta) = D G_p(\theta)$, we obtain

$$(3.27) \quad \partial_{rs} \Pr \{m \text{ tr } S_1 \Lambda \leq 2\theta\} = \frac{m}{2} (rs) E g_p(\theta),$$

$$(3.28) \quad \begin{aligned} & \partial_{rs} \partial_{tu} \Pr \{m \text{ tr } S_1 \Lambda \leq 2\theta\} \\ &= - \left\{ \frac{m}{2} (rs | tu) (E^2 + E) + \frac{m^2}{4} (rs)(tu) (E^3 - E) \right\} g_p(\theta), \end{aligned}$$

$$(3.29) \quad \begin{aligned} & \partial_{rs} \partial_{tu} \partial_{vw} \Pr \{m \text{ tr } S_1 \Lambda \leq 2\theta\} = \left\{ m(rs | tu | vw) (E^3 + E^2 + E) + \frac{m^3}{4} \right. \\ & \cdot [(rs)(tu | vw) + (tu)(rs | vw) + (vw)(rs | tu)] (E^3 - E) \\ & \left. + \frac{m^2}{8} (rs)(tu)(vw) (E^3 - 2E^2 + E) \right\} \cdot g_p(\theta), \end{aligned}$$

$$(3.30) \quad \begin{aligned} & \partial_{rs} \partial_{tu} \partial_{vw} \partial_{xy} \Pr \{m \text{ tr } S_1 \Lambda \leq 2\theta\} \\ &= - \left\{ m[(rs | tu | vw | xy) \right. \\ & + (rs | vw | xy | tu) + (rs | xy | tu | vw)] (E^4 + E^3 + E^2 + E) \\ & + \frac{m^2}{2} [(rs)(tu | vw | xy) + (xy)(tu | vw | rs) + (vw)(tu | xy | rs) \\ & + (tu)(vw | xy | rs)] (E^4 - E) + \frac{m^2}{4} [(rs | tu)(vw | xy) \\ & + (rs | vw)(tu | xy) + (rs | xy)(tu | vw)] (E^4 + E^3 - E^2 - E) \\ & + \frac{m^3}{8} [(rs)(tu)(vw | xy) + (rs)(vw)(tu | xy) \\ & + (rs)(xy)(tu | vw) + (tu)(vw)(rs | xy) + (tu)(xy)(rs | vw) \\ & + (vw)(xy)(rs | tu)] (E^4 - E^3 - E^2 + E) \\ & \left. + \frac{m^4}{16} (rs)(tu)(vw)(xy) (E^4 - 3E^3 + 3E^2 - E) \right\} g_p(\theta). \end{aligned}$$

Upon substituting (3.28) into (3.16), we obtain

$$h_1(\Lambda^{-1}) = \frac{1}{4n} \sum \sigma_{rs} \sigma_{tu} \left[2m(st | ur) \left\{ \frac{\theta^2}{\rho(\rho+1)} + \frac{\theta}{\rho} \right\} + m^2(st)(ur) \left\{ \frac{\theta^2}{\rho(\rho+1)} - \frac{\theta}{\rho} \right\} \right].$$

Now,

$$(st) = \text{tr } \Lambda_{st}^{-1} \Lambda = \frac{1}{2} \sum_{i,j} (\delta_{si} \delta_{tj} + \delta_{ti} \delta_{sj}) \sigma^{ji} = \frac{1}{2} (\sigma^{ts} + \sigma^{st}) = \sigma^{st}$$

and also,

$$(st | ur) = \text{tr } (\Lambda_{st}^{-1} \Lambda) (\Lambda_{ur}^{-1} \Lambda) = \frac{1}{2} (\sigma^{rs} \sigma^{tu} + \sigma^{su} \sigma^{tr}).$$

Hence we have

$$\sum \sigma_{rs} \sigma_{tu} (st | ur) = \frac{1}{2} p(p+1)$$

and

$$\sum \sigma_{rs} \sigma_{tu} (st)(ur) = p.$$

We also note that $2\theta = \chi^2$, $\rho = mp/2$. Therefore we finally obtain, after some simplification,

$$(3.31) \quad h_1(\Lambda^{-1}) = \frac{1}{4n} \left\{ \frac{p+m+1}{mp+2} \chi^4 + (p-m+1) \chi^2 \right\}.$$

In a similar way we substitute (3.29), (3.30), and (3.31) into (3.17) to evaluate $h_2(\Lambda^{-1})$. We note here that since $h_1(\Lambda^{-1})$ given by (3.31) is independent of Λ^{-1} , the terms involving $h_1^{(st)}(\Lambda^{-1})$ and $h_1^{(st,ur)}(\Lambda^{-1})$ in (3.17) do not appear. As before, it can be easily shown that

$$\begin{aligned} \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} (st | uv | wr) &= \frac{1}{2} p(p^2 + 3p + 4), \quad \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} (st)(uv | wr) \\ &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} (uv)(st | wr) = \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} (wr)(st | uv) = \frac{1}{2} p(p+1), \\ \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} (st)(uv)(wr) &= p, \quad \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st | ur | vx | yv) \\ &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st | vx | yv | ur) = \frac{1}{2} p(p+1)^2, \\ \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st | yv | ur | vx) &= \frac{1}{2} p(p+3), \quad \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st)(ur | vx | yv) \\ &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (yv)(ur | vx | st) = \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (vx)(ur | yv | st) \\ &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (ur)(vx | yv | st) = \frac{1}{2} p(p+1), \\ \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st | ur)(vx | yv) &= \frac{1}{2} p^2(p+1)^2, \quad \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st | vx)(ur | yv) \\ &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st | yv)(ur | vx) = \frac{1}{2} p(p+1), \\ \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st)(ur)(vx | yv) &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (vx)(yv)(st | ur) = \frac{1}{2} p^2(p+1), \\ \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st)(vx)(ur | yv) &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st)(yv)(ur | vx) \\ &= \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (ur)(vx)(st | yv) = \sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (ur)(yv)(st | vx) = p, \end{aligned}$$

and

$$\sum \sigma_{rs} \sigma_{tu} \sigma_{vw} \sigma_{xy} (st)(ur)(wx)(yv) = p^2.$$

Using these results we obtain from (3.17), after some simplification,

$$\begin{aligned} h_2(\Lambda^{-1}) = & \frac{1}{48n^2} \left[\frac{6(p-1)(p+2)(m-1)(m+2)}{(mp+2)^2(mp+4)(mp+6)} x^8 \right. \\ & + \frac{4mp^3 + 2(3m^2 + 3m + 10)p^2}{(mp+2)^2(mp+4)} x^6 \\ & + \frac{13p^2 + 24p - 11m^2 + 7}{mp+2} x^4 \\ & \left. + \{7p^2 + (-12m + 12)p + (7m^2 - 12m + 1)\} x^2 \right], \end{aligned} \quad (3.32)$$

which is independent of Λ^{-1} just as $h_1(\Lambda^{-1})$.

Now we substitute (3.31) and (3.32) into (3.13) to obtain

$$\begin{aligned} T_0^2 = & 2h(S_0) = 2\theta + 2h_1(S_0) + 2h_2(S_0) + O(n^{-3}) \\ = & x^2 + \frac{1}{2n} \left\{ \frac{p+m+1}{mp+2} x^4 + (p-m+1)x^2 \right\} \\ & + \frac{1}{24n^2} \left\{ \frac{6(p-1)(p+2)(m-1)(m+2)}{(mp+2)^2(mp+4)(mp+6)} x^8 \right. \\ & + \frac{4mp^3 + 2(3m^2 + 3m + 10)p^2 + 2(2m^3 + 3m^2 + 17m + 18)p}{(mp+2)^2(mp+4)} x^6 \\ & + \frac{13p^2 + 24p - 11m^2 + 7}{mp+2} x^4 \\ & \left. + [7p^2 + (-12m + 12)p + (7m^2 - 12m + 1)]x^2 \right\} + O(n^{-3}), \end{aligned} \quad (3.33)$$

which is the asymptotic expression of a percentage point of the T_0^2 distribution in terms of the corresponding percentage point of the χ^2 distribution with mp degrees of freedom.

If we put $m = 1$ in (3.33), we have

$$\begin{aligned} T^2 = & x^2 + \frac{1}{2n} \{x^4 + px^2\} \\ & + \frac{1}{24n^2} \{4x^6 + (13p-2)x^4 + (7p^2-4)x^2\} + O(n^{-3}), \end{aligned} \quad (3.34)$$

which is the asymptotic expression of a percentage point of the generalized Student T distribution. This result, (3.34), was previously obtained by Hotelling and Frankel [3], [4].

There is another check of (3.33) by putting $p = 1$ in the formula.³ In this case we have

$$(3.35) \quad T^2 = \chi^2 + \frac{1}{2n} \{ \chi^4 - (m-2)\chi^2 \} \\ + \frac{1}{24n^2} \{ 4\chi^6 - 11(m-2)\chi^4 + (m-2)(7m-10)\chi^2 \} + O(n^{-3}),$$

which is the correct expansion for the ordinary variance ratio F with m, n degrees of freedom in terms of χ^2 with m degrees of freedom [1].

4. Asymptotic formula for the c.d.f. of T_0^2 . Let $F(2\theta_1)$ be the c.d.f. of T_0^2 , i.e.,

$$(4.1) \quad F(2\theta_1) = \Pr \{ m \operatorname{tr} S_1 S_0^{-1} \leq 2\theta_1 \}.$$

Then, as (3.6), we can write

$$(4.2) \quad \Pr \{ m \operatorname{tr} S_1 S_0^{-1} \leq 2\theta_1 \} = \int_R \Pr \{ m \operatorname{tr} S_1 S_0^{-1} \leq 2\theta_1 | S_0 \} \Pr \{ dS_0 \} \\ = \Theta \Pr \{ m \operatorname{tr} S_1 \Lambda \leq 2\theta_1 \},$$

where Θ is given by (3.12). Upon substituting (3.28), (3.29), and (3.30) into (4.2) we obtain, after some simplification,

$$(4.3) \quad F(2\theta_1) = G_p(\theta_1) - \frac{1}{2n} \left\{ \frac{2(p+m+1)\theta_1^2}{mp+2} + (p-m+1)\theta_1 \right\} g_p(\theta_1) \\ - \frac{1}{48n^2} \left[\frac{24\{mp^3 + 2(m^2+m+4)p^2 + (m^3+2m^2+21m+20)p + 8m^2+20m+20\}\theta_1^4}{(mp+2)(mp+4)(mp+6)} \right. \\ + \frac{4\{3mp^3 - 2(3m^2-3m-4)p^2 - 3(3m^3+2m^2+11m-4)p - 40m^2-36m-4\}\theta_1^3}{(mp+2)(mp+4)} \\ + \frac{2\{3mp^3 + 2(3m^2+3m-4)p^2 - 3(3m^3-2m^2-5m+4)p - 8m^2+12m+4\}\theta_1^2}{mp+2} \\ \left. - \{3mp^3 - 2(3m^2-3m+4)p^2 + 3(m^3-2m^2+5m-4)p - 8m^2+12m+4\}\theta_1 \right] g_p(\theta_1) + O(n^{-3}),$$

³ The author is indebted to the referee for pointing out this check of (3.33).

where

$$G_p(\theta_1) = [\Gamma(p)]^{-1} \int_0^{\theta_1} t^{p-1} e^{-t} dt, \quad g_p(\theta_1) = \frac{\partial}{\partial \theta_1} G_p(\theta_1), \quad \text{and } p = mp/2.$$

(4.3) is a sort of multivariate analogue of Hartley's formula of "Studentization." In fact it can be shown that when $p = 1$, (4.3) coincides with Hartley's formula for the c.d.f. of the univariate analysis of variance F statistic. (See equation (28), p. 178, [2].)

5. Discussion of the error and remarks. In view of the methods used in Sections 3 and 4, it is rather difficult to set a bound for the error committed by omitting all terms after the first few terms in the asymptotic formula for T_0^2 (3.33) or in the asymptotic formula for the c.d.f. of T_0^2 (4.3). There is, however, a method to find lower and upper bounds to the c.d.f. of T_0^2 which is fairly good for large values of n , and they can be used to set a bound for $O(n^{-3})$, say, in the asymptotic expansion of the c.d.f. of T_0^2 .

We shall first obtain lower and upper bounds for the c.d.f. of T_0^2 . It is well known (e.g., see [7]) that the joint probability law of the characteristic roots e_1, e_2, \dots, e_s of $m S_1 S_0^{-1}$ under the null hypothesis H_0 is given by

$$(5.1) \quad P(e_1, e_2, \dots, e_s) = C(s, t, p, n) \prod_{i=1}^s e_i^{(t-s-1)/2} \left(1 + \frac{e_i}{n}\right)^{-(m+n)/2} de_i \prod_{i < j=1}^{s-1} (e_i - e_j),$$

where $0 \leq e_s \leq e_{s-1} \leq \dots \leq e_1 < \infty$, $s = \min(p, m)$, $t = \max(p, m)$, and

$$C(s, t, p, n) = \frac{\pi^{s/2}}{n^{st/2}} \prod_{i=1}^s \frac{\Gamma\{\frac{1}{2}(n+t-p+i)\}}{\Gamma\{\frac{1}{2}(t-s+i)\} \Gamma\{\frac{1}{2}(n-p+i)\} \Gamma(i/2)}.$$

The statistic T_0^2 is expressed as

$$(5.2) \quad T_0^2 = m \operatorname{tr} S_1 S_0^{-1} = \sum_{i=1}^s e_i,$$

and the c.d.f. of T_0^2 is given by

$$(5.3) \quad F(2\theta_1) = C(s, t, p, n) \int_{R_1} \dots \int \prod_{i=1}^s e_i^{(t-s-1)/2} \left(1 + \frac{e_i}{n}\right)^{-(m+n)/2} de_i \prod_{i < j=1}^{s-1} (e_i - e_j),$$

where R_1 is the domain of integration such that $0 \leq e_s \leq e_{s-1} \leq \dots \leq e_1 < \infty$ and $0 \leq \sum_{i=1}^s e_i \leq 2\theta_1$. Now for any non-negative values of e_i and n , the following inequality holds:

$$\log \left(1 + \frac{e_i}{n}\right) \leq \frac{e_i}{n}$$

for $i = 1, \dots, s$, where equality holds when $e_i = 0$ or $n \rightarrow \infty$. Hence we have

$$\prod_{i=1}^s \left(1 + \frac{e_i}{n}\right)^{-(m+n)/2} \geq \exp \left[-\frac{m+n}{2n} \sum_{i=1}^s e_i \right].$$

Therefore, the probability law (5.1) is bounded from below as follows:

$$(5.4) \quad P_1(e_1, \dots, e_s) \leq P(e_1, \dots, e_s)$$

where

$$P_1(e_1, \dots, e_s) = C(s, t, p, n) \prod_{i=1}^s e_i^{(t-s-1)/2} de_i \exp \left[-\frac{m+n}{2n} \sum_{i=1}^s e_i \right] \prod_{i < j=1}^{s-1} (e_i - e_j).$$

It must be noted here that $P_1(e_1, \dots, e_s)$ is not a probability law, although it is non-negative for all e_i such that $0 \leq e_s \leq \dots \leq e_1 < \infty$. Now integrating both sides of (5.4) in R_1 we obtain

$$(5.5) \quad F_1(2\theta_1) \leq F(2\theta_1),$$

where

$$F_1(2\theta_1) = C(s, t, p, n) \int_{R_1} \dots \int \prod_{i=1}^s e_i^{(t-s-1)/2} de_i \exp \left[-\frac{m+n}{2n} \sum_{i=1}^s e_i \right] \prod_{i < j=1}^{s-1} (e_i - e_j),$$

and also integrating both sides of (5.4) in R_2 where $0 \leq e_s \leq \dots \leq e_1 < \infty$ and $2\theta_1 \leq \sum_{i=1}^s e_i < \infty$ and subtracting each from 1, we have

$$(5.6) \quad F(2\theta_1) \leq F_2(2\theta_1),$$

where

$$F_2(2\theta_1) = 1 - C(s, t, p, n) \int_{R_2} \dots \int \prod_{i=1}^s e_i^{(t-s-1)/2} de_i \exp \left[-\frac{m+n}{2n} \sum_{i=1}^s e_i \right] \prod_{i < j=1}^{s-1} (e_i - e_j).$$

In order to evaluate $F_1(2\theta_1)$ and $F_2(2\theta_1)$, we observe that as n tends to ∞ , $T_0^2 = \sum_{i=1}^s e_i$ has a χ^2 distribution with st degrees of freedom in the limit; i.e., we have

$$(5.7) \quad K(s, t, p) \int_{R_1} \dots \int \prod_{i=1}^s e_i^{(t-s-1)/2} de_i \exp \left[-\frac{1}{2} \sum_{i=1}^s e_i \right] \prod_{i < j=1}^{s-1} (e_i - e_j) = G_{st}(\theta_1),$$

where

$$K(s, t, p) = \lim_{n \rightarrow \infty} C(s, t, p, n) = \frac{\pi^{s/2}}{2^{st/2}} \frac{1}{\prod_{i=1}^s \Gamma(\frac{1}{2}(t-s+i)) \Gamma(\frac{i}{2})}$$

and $\rho_1 = st/2$. Hence integration of (5.5) yields

$$(5.8) \quad F_1(2\theta_1) = L(s, t, p, n) G_{\rho_1} \left(\frac{m+n}{n} \theta_1 \right),$$

where

$$L(s, t, p, n) = \frac{C(s, t, p, n)}{K(s, t, p)} \left(\frac{n}{m+n} \right)^{st/2} = \left(\frac{2}{m+n} \right)^{st/2} \prod_{i=1}^s \frac{\Gamma \left(\frac{n+t-p+i}{2} \right)}{\Gamma \left(\frac{n-p+i}{2} \right)}.$$

Similarly we obtain from (5.6)

$$(5.9) \quad F_2(2\theta_1) = 1 - L(s, t, p, n) \left\{ 1 - G_{\rho_1} \left(\frac{m+n}{n} \right) \right\}.$$

Now if we write (4.3) as

$$(5.10) \quad F(2\theta_1) = a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + R_3,$$

where R_3 is the error committed by omitting all terms except the first three terms in the asymptotic series of $F(2\theta_1)$, the absolute value of R_3 has the following upper bound:

$$(5.11) \quad |R_3| \leq \max \left\{ \left| F_1(2\theta_1) - a_0 - \frac{a_1}{n} - \frac{a_2}{n^2} \right|, \left| F_2(2\theta_1) - a_0 - \frac{a_1}{n} - \frac{a_2}{n^2} \right| \right\},$$

where $F_1(2\theta_1)$ and $F_2(2\theta_1)$ are given by (5.8) and (5.9), respectively.

The actual manner in which (3.33) converges to the true value T_0^2 or in which (4.3) converges to the true value $F(2\theta_1)$ is not known, but it is hoped that the use of the first few corrective terms may result in a test which is more accurate than the χ^2 approximation, at any rate for moderately large values of n . In the case of the asymptotic formula for the c.d.f. of T_0^2 (4.3), we may judge the magnitude of the error involved in using the first few terms of the series by (5.11), which turns out to be rather small numerically when n is sufficiently large.

The author wishes to express his indebtedness to Professor Harold Hotelling for suggesting this problem and for his guidance in the preparation of this paper.

REFERENCES

- [1] J. R. GREEN, "A confidence interval for variance components," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 671-686.
- [2] H. O. HARTLEY, "Studentization, or the elimination of the standard deviation of the

- parent population from the random sample-distribution of statistics," *Biometrika*, Vol. 33 (1944), pp. 173-180.
- [3] H. HOTELLING, "A generalized T -test and measure of multivariate dispersion," *Proc. Second Berkeley Symposium on Math. Stat. and Probability*, 1950, pp. 23-41.
 - [4] H. HOTELLING AND L. R. FRANKEL, "The transformation of statistics to simplify their distribution," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 87-96.
 - [5] G. S. JAMES, "The comparison of several groups of observations when the ratios of the population variances are unknown," *Biometrika*, Vol. 38 (1951), pp. 324-329.
 - [6] G. S. JAMES, "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown," *Biometrika*, Vol. 41 (1954), pp. 19-43.
 - [7] S. N. ROY, " p -statistics and some generalizations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381-396.
 - [8] B. L. WELCH, "The generalization of 'Student's' problem when several different population variances are involved," *Biometrika*, Vol. 34 (1947), pp. 28-35.

SOME RESULTS USEFUL IN MULTIVARIATE ANALYSIS¹

By K. C. S. PILLAI²

University of North Carolina and University of Travancore

1. Summary and Introduction. In this paper a few results (of a purely mathematical nature) are obtained, which are useful for studying certain distribution problems in multivariate analysis—e.g., those relating to the characteristic roots of a determinantal equation ([1], [2], [5]). In particular, the results are shown to be readily applicable to the moment problems of the sum of the roots and the distributions of the extreme roots. Most of the results given are in the form of certain recursion formulae for reducing special types of k -th order Vandermonde determinants in terms of those of orders $(k - 1)$ and $(k - 2)$. The applications of these results are given by S. N. Roy [6] and the present author [3].

2. Vandermonde's determinant. Let us first consider a type of determinant (due to Vandermonde) which plays an important role in the development of this paper. Denote by V_0 the Vandermonde's determinant of the form

$$(2.1) \quad V_0 = \begin{vmatrix} X_k^{k-1} & X_k^{k-2} & \cdots & X_k & 1 \\ X_{k-1}^{k-1} & X_{k-1}^{k-2} & \cdots & X_{k-1} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_2^{k-1} & X_2^{k-2} & \cdots & X_2 & 1 \\ X_1^{k-1} & X_1^{k-2} & \cdots & X_1 & 1 \end{vmatrix},$$

where X_1, X_2, \dots, X_k are k variables. The determinant can be shown to be equal to the expression

$$(2.2) \quad V_0 = \prod_{i>j} (X_i - X_j),$$

where \prod denotes the product over the k variables. The determinant V_0 has several interesting properties, of which the following will be used in this paper.

Property 1. If each of the indices of the first j columns of V_0 is increased by unity, the resulting determinant

$$(2.3) \quad V_j \text{ (say)} = (\sum X_1 X_2 \cdots X_j) V_0,$$

where $\sum X_1 X_2 \cdots X_j$ denotes the j -th elementary symmetric function in k variables X_1, X_2, \dots, X_k .

Received September 28, 1955; revised July 9, 1956.

¹ Part of a doctoral thesis submitted to the Department of Statistics, University of North Carolina, Chapel Hill; work done under the sponsorship of the Ford Foundation.

² Now with United Nations, New York.

3. A special function and the corresponding determinant. Let us consider the integral over the domain $0 < x_1 \leq x_2 \leq \dots \leq x_k \leq x < 1$, of a function given by

$$(3.1) \quad f(x_1, x_2, \dots, x_k) = \prod_{i=1}^k \{x_i^q (1 - x_i)^r e^{tx_i}\} \prod_{i>j} (x_i - x_j),$$

where $q, r > -1$ and the t is independent of the x 's. It is obvious that, in view of (2.2), $\prod_{i>j} (x_i - x_j)$ can be thrown into the form of a determinant as in (2.1). Multiply the i -th row of this determinant by $x_{k-i+1}^q (1 - x_{k-i+1})^r e^{tx_{k-i+1}}$ ($i = 1, 2, \dots, k$), and integrate between appropriate limits, each term of the determinant with respect to the variable it involves. Then the integral of the function $f(x_1, x_2, \dots, x_k)$ given by (3.1) takes the form

$$(3.2) \quad \begin{vmatrix} \int_0^x x_k^{q+k-1} (1 - x_k)^r e^{tx_k} dx_k & \dots & \int_0^x x_k^q (1 - x_k)^r e^{tx_k} dx_k \\ \dots & \dots & \dots \\ \int_0^{x_2} x_1^{q+k-1} (1 - x_1)^r e^{tx_1} dx_1 & \dots & \int_0^{x_2} x_1^q (1 - x_1)^r e^{tx_1} dx_1 \end{vmatrix}.$$

It has to be remembered that in expanding the determinant (3.2), the order of integration must not be changed, and hence we shall call it a "pseudo-determinant."

Now let q_1, q_2, \dots, q_k be real numbers greater than -1 . Let us denote by $U(x; q_k, r; \dots; q_1, r; t)$ the pseudo-determinant

$$(3.3) \quad \begin{vmatrix} \int_0^x x_k^{q_k} (1 - x_k)^r e^{tx_k} dx_k & \dots & \int_0^x x_k^{q_1} (1 - x_k)^r e^{tx_k} dx_k \\ \dots & \dots & \dots \\ \int_0^{x_2} x_1^{q_k} (1 - x_1)^r e^{tx_1} dx_1 & \dots & \int_0^{x_2} x_1^{q_1} (1 - x_1)^r e^{tx_1} dx_1 \end{vmatrix}.$$

More generally, if we replace r in the j -th column of (3.3) by

$$r_{k-j+1} (j = 1, 2, \dots, k),$$

the resulting pseudo-determinant will be denoted by

$$(3.4) \quad U(x; q_k, r_k; q_{k-1}, r_{k-1}; \dots; q_1, r_1; t),$$

or more explicitly by

$$(3.5) \quad U \left\{ x; \begin{pmatrix} q_k, r_k & \dots & q_1, r_1 \\ & \dots & \\ & & \dots \\ & & & q_k, r_k & \dots & q_1, r_1 \end{pmatrix}; t \right\}.$$

Further, in the pseudo-determinant (3.3), if the indices of the i -th row alone are different from those of the rest, we denote the resulting determinant by

$$(3.6) \quad U(x; q_k'', r''; \cdots; q_1'', r''; t)^{(i)},$$

where $q_k'', r'', \cdots, q_1'', r''$ denote the indices of the i -th row.

Since the integral of $f(x_1, x_2, \cdots, x_k)$ involves integrals of the type

$$(3.7) \quad I(x'; q, r; F; t) = \int_0^{x'} x^q (1-x)^r F(x) e^{tx} dx,$$

where $F(x)$ is a function of x such that the integral in (3.7) exists, let us first consider the integral (3.7). If $F(x)$ is of the form

$$(3.8) \quad \int_0^{x_k} x_{k-1}^{q_{k-1}-1} (1-x_{k-1})^r e^{tx_{k-1}} dx_{k-1} \cdots \int_0^{x_1} x_1^{q_1-1} (1-x_1)^r e^{tx_1} dx_1,$$

the integral (3.7) may be denoted by

$$(3.9) \quad I(x'; q, r; q_{k-1}, r; \cdots; q_1, r; t).$$

Now consider $I(x'; q, r; F; t)$. Integrating (3.7) by parts we obtain the result stated in the following lemma:

LEMMA 1.

$$(3.10) \quad \begin{aligned} I(x'; q, r; F; t) &= (q+r+1)^{-1} \{-I_0(x'; q, r+1; F; t) \\ &+ I(x'; q, r+1; F'; t) + qI(x'; q-1, r; F; t) \\ &+ tI(x'; q, r+1; F; t)\}, \end{aligned}$$

where

$$I_0(x'; q, r+1; F; t) = x^q (1-x)^{r+1} F(x) e^{tx} \Big|_0^{x'}, \quad F' = \frac{dF(x)}{dx}.$$

It may be noted that the right-hand side of (3.10) has been obtained by integrating $(1-x)^{r+q}$ and differentiating the product of $x^q/(1-x)^q$, $F(x)$ and e^{tx} , treating this product as the u term in $\int u dv$. Using Lemma 1, let us consider the integration of the function in (3.3) when $k=2$.

THEOREM 1. *The pseudo-determinant*

$$(3.11) \quad \begin{aligned} U(x; q_2, r; q_1, r; t) &= (q_2+r+1)^{-1} \{-I_0(x; q_2, r+1; q_1, r; t) \\ &+ 2I(x; q_2+q_1, 2r+1, 2t) + q_2 U(x; q_2-1, r; q_1, r; t) \\ &+ tU(x; q_2, r+1; q_1, r; t)\}. \end{aligned}$$

PROOF. First, note that $U(x; q_2, r; q_1, r; t) = I(x; q_2, r; q_1, r; t) - I(x; q_1, r; q_2, r; t)$. Integrate the latter integrals by parts using Lemma 1 so as to reduce the index q_2 in each case by unity. The sum of all the terms thus obtained after integration gives the right-hand side of (3.11). For a more detailed proof of the theorem, the reader is referred to [3].

In (3.11) the last pseudo-determinant can further be shown to be equal to the difference of two others given by

$$(3.12) \quad U(x; q_2, r+1; q_1, r; t) = U(x; q_2, r; q_1, r; t) - U(x; q_2+1, r; q_1, r; t).$$

For integration in the general case of the function contained in the pseudo-determinant (3.3), some more results have to be used. These results are stated as lemmas in the following section. For the detailed proofs of these lemmas the reader is referred to [3].

4. Certain properties of I -functions. This section is devoted to the statement of two lemmas which will be used in the next section.

LEMMA 2. If (q'_k, \dots, q'_1) denotes any permutation of (q_k, \dots, q_1) , then

$$(4.1) \quad \sum I(x; q'_k, r; \dots; q'_1, r; t) = \prod_{j=1}^k I(x; q_j, r, t),$$

where the summation \sum extends over all possible permutations.

LEMMA 3. If $U(x; q_k'', r'', t''; \dots; q_1'', r'', t'')^{(i)}$ denotes the pseudo-determinant in (3.6) with t'' for the index of the i -th row instead of t , which is the index everywhere else, then

$$(4.2) \quad \begin{aligned} & \sum_{i=1}^k (-1)^{i-1} U(x; q_k'', r'', t''; \dots; q_1'', r'', t'')^{(i)} \\ &= \sum_{j=k}^1 (-1)^{k-j} I(x; q_j'', r'', t'') U(x; q_k, r; \dots; q_{j+1}, r; q_{j-1}, r; \dots; q_1, r; t). \end{aligned}$$

5. Pseudo determinant of order 3. In this section we shall prove the following theorem:

THEOREM 2. The pseudo-determinant

$$(5.1) \quad \begin{aligned} & U(x; q_3, r; q_2, r; q_1, r; t) \\ &= (q_3 + r + 1)^{-1} \{ -I_0(x; q_3, r + 1; t) U(x; q_2, r; q_1, r; t) \\ & \quad + 2I(x; q_3 + q_2, 2r + 1; 2t) I(x; q_1, r, t) \\ & \quad - 2I(x; q_3 + q_1, 2r + 1; 2t) I(x; q_2, r, t) \\ & \quad + q_3 U(x; q_3 - 1, r; q_2, r; q_1, r; t) \\ & \quad + t U(x; q_3, r + 1; q_2, r; q_1, r; t) \}. \end{aligned}$$

PROOF. Expand the pseudo-determinant $U(x; q_3, r; q_2, r; q_1, r; t)$ as follows:

$$(5.2) \quad \begin{aligned} & U \left\{ x; \begin{pmatrix} q_3, r \\ & q_2, r & q_1, r \\ & q_2, r & q_1, r \end{pmatrix}; t \right\} + U \left\{ x; \begin{pmatrix} & q_2, r & q_1, r \\ q_3, r & & \\ & q_2, r & q_1, r \end{pmatrix}; t \right\} \\ & \quad + U \left\{ x; \begin{pmatrix} & q_2, r & q_1, r \\ & q_2, r & q_1, r \\ q_3, r & & \end{pmatrix}; t \right\}. \end{aligned}$$

It has to be understood that in each of the pseudo-determinants in (5.2), there are no elements in the positions left blank. Each component in (5.2) stands for the product of an element of the first column and its cofactor in the third-order pseudo-determinant $U(x; q_3, r; q_2, r; q_1, r; t)$. Since the order of integration must not be changed, the product is not written explicitly. Now using Lemma 1, integrate by parts the first pseudo-determinant in (5.2) with respect to x_3 , the second with respect to x_2 , and the third with respect to x_1 . Add the expressions obtained corresponding to each of the four terms on the right-hand side of (3.10). This yields

$$(5.3) \quad (q_3 + r + 1)^{-1}(A^{(3)} + B^{(3)} + q_3 C^{(3)} + t D^{(3)}),$$

where

$$(5.4) \quad A^{(3)} = -I_0(x; q_3, r + 1; t)U(x; q_2, r; q_1, r; t);$$

$$(5.5) \quad B^{(3)} = 2 \sum_{i=1}^2 (-1)^{i-1} U(x; q_3 + q_2, 2r + 1, 2t; q_3 + q_1, 2r + 1, 2t)^{(i)};$$

$$(5.6) \quad C^{(3)} = U(x; q_3 - 1, r; q_2, r; q_1, r; t);$$

and

$$(5.7) \quad D^{(3)} = U(x; q_3, r + 1; q_2, r; q_1, r; t).$$

Now apply Lemma 3 to the right-hand side of (5.5) with $k = 2$; we at once get the result (5.1).

6. Pseudo determinant of order k . We generalize the results of Theorem 2 in the following theorem:

THEOREM 3. *The pseudo-determinant*

$$(6.1) \quad U(x; q_k, r; q_{k-1}, r; \dots; q_1, r; t) \\ = (q_k + r + 1)^{-1}(A^{(k)} + B^{(k)} + q_k C^{(k)} + t D^{(k)}),$$

where

$$(6.2) \quad A^{(k)} = -I_0(x; q_k, r + 1; t)U(x; q_{k-1}, r; \dots; q_1, r; t);$$

$$(6.3) \quad B^{(k)} = 2 \sum_{j=k-1}^1 (-1)^{k-j-1} I(x; q_k + q_j, 2r + 1, 2t) \\ \cdot U(x; q_{k-1}, r; \dots; q_{j+1}, r; q_{j-1}, r; \dots; q_1, r; t);$$

$$(6.4) \quad C^{(k)} = U(x; q_k - 1, r; q_{k-1}, r; \dots; q_1, r; t);$$

and

$$(6.5) \quad D^{(k)} = U(x; q_k, r + 1; q_{k-1}, r; \dots; q_1, r; t).$$

The proof of this theorem follows step by step that of Theorem 2.

It may be noted that the pseudo-determinant in (6.5) can be expressed as a difference of two others as given below:

$$(6.6) \quad U(x; q_k, r + 1; q_{k-1}, r; \dots; q_1, r; t) = U(x; q_k, r; \dots; q_1, r; t) \\ - U(x; q_k + 1, r; q_{k-1}, r; \dots; q_1, r; t).$$

Further note that if $q_j = q_{j-1} + 1$ in Theorem 3, the pseudo-determinant (6.4) vanishes.

Now we state below another theorem which can be proved by employing techniques similar to those used to prove Theorem 3.

THEOREM 4. If $W(y; a_k, b; a_{k-1}, b; \dots; a_1, b; -t)$ denotes the pseudo-determinant

$$(6.7) \quad \begin{vmatrix} \int_0^y y_k^{a_k} e^{-ty_k} / (1 + y_k)^b dy_k & \dots & \int_0^y y_k^{a_1} e^{-ty_k} / (1 + y_k)^b dy_k \\ \dots & \dots & \dots \\ \int_0^{y_2} y_1^{a_k} e^{-ty_1} / (1 + y_1)^b dy_1 & \dots & \int_0^{y_2} y_1^{a_1} e^{-ty_1} / (1 + y_1)^b dy_1 \end{vmatrix}$$

where $a_i > -1, b > a_i + 1, (i = 1, 2, \dots, k)$ and

$$0 < y_1 \leq y_2 \leq \dots \leq y_k \leq y < \infty;$$

then

$$(6.8) \quad \begin{aligned} W(y; a_k, b; a_{k-1}, b; \dots; a_1, b; -t) \\ = (b - a_k - 1)^{-1} (P^{(k)} + Q^{(k)} + a_k R^{(k)} - t S^{(k)}), \end{aligned}$$

where

$$(6.9) \quad P^{(k)} = -F_0(y; a_k, b - 1, -t) W(y; a_{k-1}, b; \dots; a_1, b; -t);$$

$$(6.10) \quad \begin{aligned} Q^{(k)} = 2 \sum_{j=1}^k F(y; a_k + a_j, 2b - 1, -2t) \\ \cdot W(y; a_{k-1}, b; \dots; a_{j+1}, b; a_{j-1}, b; \dots; a_1, b; -t), \end{aligned}$$

$$(6.11) \quad R^{(k)} = W(y; a_k - 1, b; a_{k-1}, b; \dots; a_1, b; -t),$$

and

$$(6.12) \quad S^{(k)} = W(y; a_k, b - 1; a_{k-1}, b; \dots; a_1, b; -t).$$

The pseudo determinant (6.12) can be expressed as the sum of two others, as follows:

$$(6.13) \quad S^{(k)} = W(y; a_k, b; \dots; a_1, b; -t) + W(y; a_k + 1, b; \dots; a_1, b; -t).$$

7. Applications to multivariate analysis. The results given by Theorems 3 and 4 are useful for certain distribution problems in multivariate analysis. Consider the well-known distribution of the non-zero roots ($0 < \theta_1 \leq \theta_2 \leq \dots \leq \theta_s < 1$; $s \leq p$, the number of variates) of a determinantal equation in multivariate analysis given by R. A. Fisher [1], P. L. Hsu [2] and S. N. Roy [5]. It can be written in the form

$$(7.1) \quad p(\theta_1, \dots, \theta_s) = C(s, m, n) \prod_{i=1}^s \theta_i^m (1 - \theta_i)^n \prod_{i>j} (\theta_i - \theta_j) \\ 0 < \theta_1 \leq \dots \leq \theta_s < 1,$$

where

$$C(s, m, n) = \pi^{s/2} \prod_{i=1}^s \Gamma\{(2m + 2n + s + i + 2)/2\} / \prod_{i=1}^s \Gamma\{(2m + i + 1)/2\} \cdot \Gamma\{(2n + i + 1)/2\} \Gamma(i/2).$$

For the interpretation of m and n , see [4].

Now let $V^{(s)} = \sum_{i=1}^s \theta_i$. Consider the moment generating function of $V^{(s)}$ given by $E\{\exp tV^{(s)}\}$, where E denotes mathematical expectation. It is easy to see that, in (3.1), if we put

$$(7.3) \quad q = m, \quad r = n, \quad k = s, \quad x_i = \theta_i, \quad \text{and} \quad x = 1,$$

multiply the resulting expression by $C(s, m, n)$ given in (7.2), and integrate with respect to the θ_i 's over the domain $0 < \theta_1 \leq \dots \leq \theta_s < 1$, we at once obtain $E\{\exp tV^{(s)}\}$. In other words, $E\{\exp tV^{(s)}\}$ is obtained from the pseudo-determinant (3.2) after substitutions (7.3) and multiplication by $C(s, m, n)$. Now apply Theorem 3 to $E\{\exp tV^{(s)}\}$. We obtain

$$(7.4) \quad (m + n + s)E(e^{tV^{(s)}}) = 2C(s, m, n) \sum_{j=1}^{s-1} (-1)^{s-j-1} \cdot \{I(1; 2m + s + j - 2, 2n + 1, 2t) \\ \times U(1; m + s - 2, n; \dots; m + j, n; m + j - 2, n; \dots; m, n; t) \\ + tC(s, m, n)U(1; m + s - 1, n + 1; m + s - 2, n; \dots; m, n; t)\}.$$

The simplification here resulted from the fact that the A term (since $x = 1$) and (since $q_k = q_{k-1} + 1$) the C term of Theorem 3 both vanish. Now in view of the result (6.6),

$$(7.5) \quad U(1; m + s - 1, n + 1; m + s - 2, n; \dots; m, n; t) \\ = (1/C(s, m, n))E\{\exp tV^{(s)}\} \\ - U(1; m + s, n; m + s - 2, n; \dots; m, n; t).$$

Further, using property 1 given in (2.3)

$$(7.6) \quad U(1; m + s - 2, n; \dots; m + j, n; m + j - 2, n; \dots; m, n; t) \\ = (1/C(s - 2, m, n))E\{(\sum \theta_1 \dots \theta_{s-j-1})e^{tV^{(s-2)}}\},$$

where $\sum \theta_1 \dots \theta_{s-j-1}$ denotes the $(s - j - 1)$ -th elementary symmetric function in $(s - 2)$ variables $\theta_1 \dots \theta_{s-2}$. Again using the same property in (7.5),

$$(7.7) \quad U(1; m + s, n; m + s - 2, n; \dots; m, n; t) = (1/C(s, m, n))E(V^{(s)}e^{tV^{(s)}}).$$

Now make use of (7.5) to (7.7) in (7.4) and we get

$$(7.8) \quad (m+n+s-t)E(e^{tV^{(s)}}) + tE(V^{(s)}e^{tV^{(s)}}) = \{2C(s, m, n)/C(s-2, m, n)\} \\ \sum_{j=1}^{s-1} (-1)^{s-j-1} I(1; 2m+s+j-2, 2n+1, 2t) E\{(\sum \theta_1 \cdots \theta_{s-j-1}) e^{tV^{(s-2)}}\}.$$

To illustrate the use of (7.8), let us put $s = 2$. This yields

$$(7.9) \quad (m+n-t+2)E(e^{tV^{(2)}}) + tE(V^{(2)}e^{tV^{(2)}}) \\ = 2C(2, m, n)I(1; 2m+1, 2n+1, 2t).$$

Noting that $I(1; 2m+1, 2n+1, 2t)$ is a confluent hypergeometric function which can be expanded as a power series in $2t$, and that $\exp tV^{(2)}$ also can be expanded as a power series in t , equating the coefficients of like powers of t on both sides of (7.9) yields

$$(7.10) \quad (m+n+i+2)\mu_i'^{(2)} - i\mu_{i-1}'^{(2)} = 2^i(2m+2) \cdots (2m+i+1) \\ (m+n+2)/(2m+2n+4) \cdots (2m+2n+i+3) \quad (i = 1, 2, \dots),$$

where $\mu_i'^{(2)}$ denotes the i -th raw moment for 2 roots. After successive substitutions of lower order moments given by the respective recurrence relations (7.10), we get

$$(7.11) \quad \mu_i'^{(2)} = \frac{(m+n+2)\Gamma(i+1)\Gamma(2m+2n+4)}{\Gamma(2m+2)\Gamma(m+n+3+i)} \sum_{j=1}^{i+1} 2^{i-j+1} \\ \frac{\Gamma(2m+i-j+3)\Gamma(m+n+i-j+3)}{\Gamma(2m+2n+i-j+5)\Gamma(i-j+2)}.$$

Computations of a similar nature with $s = 3$ and 4 in (7.9) and further evaluation of the central moments have yielded the following results [3]:

$$(7.12) \quad \mu_1^{(s)} = s(2m+s+1)/2(m+n+s+1) \quad (s = 1, 2, 3, 4),$$

$$(7.13) \quad \mu_2^{(s)} = s(2m+s+1)(2n+s+1) \\ (2m+2n+s+2)/4(m+n+s+1)^2 \\ (m+n+s+2)(2m+2n+2s+1) \quad (i = 1, 2, 3, 4),$$

and

$$(7.14) \quad \mu_3^{(s)} = s(n-m)(2m+s+1)(2n+s+1) \\ (m+n+1)(2m+2n+s+2)/d,$$

where

$$d = (m+n+s+1)^3(m+n+s+2)(m+n+s+3) \\ (2m+2n+2s)(2m+2n+2s+1), \quad (i = 1, 2, 3, 4).$$

For the corresponding $\mu_4^{(s)}$, the reader is referred to [3].

In addition to the usefulness of Theorem 3 in studying the moments of the

sum of the roots as outlined above, this theorem is also useful for evaluating the cumulative distribution function of the largest root, θ_s , of the determinantal equation. For the latter purpose, in Theorem 3, multiply

$$U(x; q_k, r; q_{k-1}, r; \dots; q_1, r; t)$$

by $C(s, m, n)$ given in (7.2), after making the following substitutions:

$$(7.15) \quad q_j = m + j - 1, \quad r = n, \quad k = s, \quad x_i = \theta_i, \quad \text{and} \quad t = 0.$$

In this case, the C term of Theorem 3 alone vanishes. By means of Theorem 3, we reduce the cumulative distribution function involving a pseudo-determinant of order s in terms of those of orders $(s-1)$ and $(s-2)$. Since it has been shown [3] that the cdf of the smallest root can be obtained from that of the largest, Theorem 3 is thus useful in obtaining the cdf of either of these roots.

Again if we wish to study the moments of the criterion

$$U^{(s)} = \sum_{i=1}^s \theta_i / (1 - \theta_i),$$

by using Theorem 4, we will arrive at the following result:

$$(7.16) \quad \begin{aligned} & (n' - m - s + t)E(e^{-tU^{(s)}}) + tE(U^{(s)}e^{-tU^{(s)}}) \\ &= \{2k(s, m, n')/K(s-2, m, n')\} \sum_{j=1}^{s-1} (-1)^{s-j-1} \\ & \quad F(\infty; 2m + s + j - 2, 2n - 1, -2t) \\ & \quad E\{(\sum \lambda_1 \dots \lambda_{s-j-1})e^{-tU^{(s)}}\}, \end{aligned}$$

where $\lambda_i = \theta_i/(1 - \theta_i)$, which transformation in (7.1) gives $K(s, m, n')$ from $C(s, m, n)$ and $n' = m + n + s + 1$.

For a detailed study of these applications in multivariate analysis, the reader is referred to [3].

8. Acknowledgments. I wish to acknowledge my indebtedness to Professor S. N. Roy for his kind advice and criticism in the course of my investigations. I am also indebted to Professor H. Hotelling for his active interest and valuable suggestions in the preparation of this paper.

REFERENCES

- [1] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eug.*, Vol. 9 (1939), pp. 238-249.
- [2] P. L. HSU, "On the distribution of the roots of certain determinantal equations," *Ann. Eug.*, Vol. 9 (1939), pp. 250-258.
- [3] K. C. S. PILLAI, "On some distribution problems in multivariate analysis," Mimeograph Series No. 88, Institute of Statistics, University of North Carolina, 1954.
- [4] K. C. S. PILLAI, "Some new test criteria in multivariate analysis," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 117-121.
- [5] S. N. ROY, "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381-396.
- [6] S. N. ROY, "The individual sampling distributions of the maximum, minimum and any intermediate of p-statistics on the null hypothesis," *Sankhya*, Vol. 7 (1945), pp. 133-158.

THE NONEXISTENCE OF CERTAIN STATISTICAL PROCEDURES IN NONPARAMETRIC PROBLEMS¹

BY R. R. BAHADUR AND LEONARD J. SAVAGE

The University of Chicago

1. Introduction. It seems plausible that if the population distribution of a real random variable is entirely unknown, then a sample from the population can yield little or no information about the tails of the distribution, even if the sample is obtained according to a sequential procedure. This paper gives evidence supporting and clarifying this proposition.

The paper treats in some detail problems of inference concerning the population mean μ . It is shown that there is neither an effective test of the hypothesis that $\mu = 0$, nor an effective confidence interval for μ , nor an effective point estimate of μ . These conclusions concerning μ flow from the fact that μ is sensitive to the tails of the population distribution; parallel conclusions hold for other sensitive parameters, and they can be established by the same methods as are here used for μ .

It is also shown that there exists no confidence band for the population distribution function such that the upper and lower limits of the band are themselves distribution functions; that is, no confidence band fits very well.

2. Theorems. Let \mathfrak{F} be a given set of distribution functions F, G, \dots of a real variable. Some of the theorems to be proved would be of interest even if \mathfrak{F} were required to be the class of all distributions or perhaps all distributions F with finite mean μ_F . But it is helpful to recognize that the proofs require only that \mathfrak{F} have a certain richness. Specifically, Theorem 1 and Corollaries 1 through 4 depend on the following three hypotheses:

- (i) For every $F \in \mathfrak{F}$, $\mu_F = \int_{-\infty}^{\infty} z dF$ exists and is finite.
- (ii) For every real m , there is an $F \in \mathfrak{F}$ with $\mu_F = m$.
- (iii) \mathfrak{F} is convex; that is, if $F \in \mathfrak{F}$, $G \in \mathfrak{F}$, π is a positive fraction, and $H = \pi F + (1 - \pi)G$, then $H \in \mathfrak{F}$.

Theorem 2 depends on hypotheses (iii) and the following:

- (iv) \mathfrak{F} is closed under translation; that is, if $F \in \mathfrak{F}$, and $G(z) = F(z - h)$ for all z and some h , then $G \in \mathfrak{F}$.
- (v) \mathfrak{F} is nonvacuous.

Some obvious examples of sets satisfying all four conditions are the sets of all distribution functions F such that μ_F is finite; the points of increase of F are a bounded set, or are a finite set; F is absolutely continuous and dF/dz vanishes

Received October 11, 1955.

¹ Research sponsored in part by the Office of Naval Research.

outside of a bounded interval; as z approaches ∞ , $1 - F(z) + F(-z) = O(z^{-r})$ for an $r > 1$.

Since the theorems to be proved are theorems of nonexistence, it is appropriate that they be stated and proved for mixed (i.e., randomized) procedures—sampling, estimating, testing. They are, of course, true a fortiori for the smaller class of pure procedures. The technique of working with mixed procedures is presented in detail in certain publications, for example [1] and [2]. We feel free, therefore, to handle mixed procedures rather informally, to save space and tedium.

Let X_1, X_2, \dots denote an infinite sequence of independent random variables, each distributed according to F ; that is, $\Pr(X_i \leq z) = F(z)$. Suppose that a (randomized, sequential) sampling procedure is given, that is, a set of rules for observing X_1, X_2, \dots one by one up to a certain stage N such that at each stage the decision whether to continue depends (randomly) on the observed values in hand at that stage. The given procedure, which will remain fixed throughout the discussion, is naturally assumed to be closed, that is,

$$(1) \quad P_F(N < \infty) = 1$$

for each $F \in \mathfrak{F}$. Except for this condition, the sampling procedure is arbitrary.

Denote the total outcome of the sampling procedure, regarded as a random variable, by V , that is, $V = (X_1, X_2, \dots, X_N)$. As already exemplified in (1), for any event A defined on the sample space of V , $P_F(A)$ will denote the probability of A when F obtains, that is to say, when each X_i is distributed according to F . If φ is a real valued function of V , $E_F[\varphi]$ will denote the expected value of φ (if it exists) when F obtains.

For any real number m , let \mathfrak{F}_m denote the set of all $F \in \mathfrak{F}$ with $\mu_F = m$.

THEOREM 1. *For each bounded real valued function φ on the sample space of V , $\inf_{F \in \mathfrak{F}_m} E_F[\varphi]$ and $\sup_{F \in \mathfrak{F}_m} E_F[\varphi]$ are independent of m .*

The proofs of this theorem and of Theorem 2 below are postponed to the next section. Theorem 1 states, in effect, that even if μ_F is known to equal one of two given values m_1 and m_2 , the sample V cannot provide effective discrimination between the two hypothetical values. The following Corollaries 1 through 4 exploit the close relations between discrimination, testing, and estimation to make explicit some consequences of Theorem 1 in problems of inference concerning μ_F . As was mentioned in the introduction, analogues of Theorem 1 (and therewith of Corollaries 1 through 4) are valid for parameters other than the mean, and these analogues can be proved by the same method as is used in the next section to prove Theorem 1.

Let H be the hypothesis that $\mu_F = 0$ (i.e., $F \in \mathfrak{F}_0$). For any test t , let $\beta_F(t)$ denote the probability of rejecting H in using t when F obtains, in short, the power function of t . Call t a somewhere unbiased level- α test if $\beta_F \leq \alpha$ for $F \in \mathfrak{F}_0$ and, for some m different from zero, $\beta_F \geq \alpha$ for $F \in \mathfrak{F}_m$. Call t a similar level- α test if $\beta_F = \alpha$ for each $F \in \mathfrak{F}_0$.

Taking $\varphi(V)$ to be the probability prescribed by t of rejecting H on observation of V yields this corollary.

COROLLARY 1. *If t is a somewhere unbiased level- α test of H , or a similar level- α test of H , then $\beta_F(t) = \alpha$ for all $F \in \mathfrak{F}$.*

Corollary 1 asserts the failure, in certain senses, of all tests of the value of μ , assuming that μ exists. It would be interesting to know whether, in comparable nonparametric situations, tests of the *existence* of μ are equally unsuccessful. To be precise, suppose for example that \mathfrak{F} is the set of all distribution functions. Let \mathfrak{F}^* be the subset of \mathfrak{F} on which μ_F exists finitely, and let H^* denote the hypothesis that F is in \mathfrak{F}^* . Then, does Corollary 1 hold with H replaced by H^* and "somewhere unbiased" replaced by "unbiased"?

Next, let I be a confidence set for μ_F , that is, I is a (randomized) function of V , that has Borel subsets of the real line for its values. For any real m , let $C[m]$ denote the event that I covers m .

COROLLARY 2. *If $P_F(C[\mu_F]) \geq 1 - \alpha$ for all $F \in \mathfrak{F}$, then $P_F(C[m]) \geq 1 - \alpha$ for all m and all $F \in \mathfrak{F}$.*

PROOF. For each m , let $p_m(V)$ be the conditional probability of $C[m]$ given V , $0 \leq p \leq 1$. Consider a fixed m . By hypothesis, $E_F[p_m] \geq 1 - \alpha$ for $F \in \mathfrak{F}_m$. Hence, $P_F(C[m]) = E_F[p_m] \geq 1 - \alpha$ for all $F \in \mathfrak{F}$, by Theorem 1. Since m is arbitrary, the corollary is proved.

COROLLARY 3. *Suppose that there exists at least one $F \in \mathfrak{F}$ such that $P_F(I \text{ is a set bounded from below}) = 1$. Then $\inf_{F \in \mathfrak{F}} \{P_F(C[\mu_F])\} = 0$.*

PROOF. For each $n = 1, 2, \dots$, let B_n denote the event that I is contained in the interval $[-n, \infty)$, and let \bar{B}_n denote the complement of B_n . For each n , let $q_n(V)$ denote the probability of B_n given V ; $0 \leq q_n \leq q_{n+1} \leq 1$.

Now let F be a distribution in \mathfrak{F} such that I is bounded from below with probability 1 when F obtains. By Lebesgue's theorem for monotone sequences, $E_F[\lim_n q_n] = \lim_n E_F[q_n] = \lim_n P_F(B_n) = P_F(I \text{ is bounded from below}) = 1$. Consequently, $\lim_n q_n(V) = 1$ except on a set of points V of P_F -measure zero. Since, for any $m < -n$, $p_m(V) = \Pr(m \in I | V) \leq \Pr(\bar{B}_n | V) = 1 - q_n(V)$, it follows that, except on a P_F -null set,

$$(2) \quad \lim_{m \rightarrow -\infty} p_m(V) = 0.$$

Since $P_F(C[m]) = E_F[p_m]$ for all m , it follows from (2), by Lebesgue's theorem for boundedly convergent sequences, that

$$(3) \quad \lim_{m \rightarrow -\infty} P_F(C[m]) = 0.$$

Now, Corollary 2 states in effect that $\inf_{F \in \mathfrak{F}} \{P_F(C[\mu_F])\} = \inf_{F \in \mathfrak{F}} \{P_F(C[m])\}$. It follows from (3) that the common value of these infima is zero. This completes the proof.

Of course, "set bounded from above," and, a fortiori, "bounded set" can be substituted for "set bounded from below" in the statement of Corollary 3. But the following example shows that it would not be enough to say "set bounded from above or from below." For all V , let $I = (-\infty, 0]$ with probability $\frac{1}{2}$ and $I = (0, \infty)$ with probability $\frac{1}{2}$; then $P_F(C[m]) = \frac{1}{2}$ for all m and all F .

Next, consider the problem of constructing a suitable point estimator for μ_F . Let M be an estimator, that is, a real valued (randomized) function of V . Suppose that when F obtains, the expected loss in using M is $E_F[L(M - \mu_F)] = r_F(M)$, where $L(m)$ is bounded from below and $\lim_{m \rightarrow \infty} L(m) = \infty$ or $\lim_{m \rightarrow -\infty} L(m) = \infty$ (e.g., $L(m) = |m|$, $L(m) = m^2$, $L(m) = (2 + \sin m)e^m$).

Let $\rho(F)$ be a real valued functional on \mathfrak{F} . Say that ρ is uncontrollable (from above) if there exists no real valued (randomized) function of V , say S , such that $\inf_{F \in \mathfrak{F}} \{P_F(\rho(F) < S)\} > 0$.

The following corollary shows that there is no estimator M for which the expected loss $r_F(M)$ is bounded in F , nor even one for which the sample gives any clue as to the possible expected loss.

COROLLARY 4. *For any estimator M , $r_F(M)$ is uncontrollable.*

PROOF. There is no loss in generality in assuming that $\lim_{m \rightarrow \infty} L(m) = \infty$. Replacing $L(m)$ by $L(m) - \inf_a L(a)$, there is also no loss in assuming L non-negative, with $\inf_m L(m) = 0$. Consider a fixed estimator M . Write $L_F = L(M - \mu_F)$. Since $L_F \geq 0$, it is easily seen (a la Tchebycheff) by considering the cases $r_F = 0$, $0 < r_F < \infty$, and $r_F = \infty$ separately that $P_F(L_F \leq \alpha r_F) \geq 1 - (1/\alpha)$ for all $\alpha > 0$ and all F .

Suppose, contrary to Corollary 4, that there exists a random variable S with distribution determined by V , and a positive constant β , such that $P_F(r_F < S) \geq \beta$ for all $F \in \mathfrak{F}$. There is no loss of generality in assuming that S is always positive. Choose and fix an $\alpha > 0$ such that $\beta - (1/\alpha) > 0$. Let $Y = \sup \{m: L(m) \leq \alpha S\}$ and define I to be the random interval $[M - Y, \infty)$. Then $P_F(I$ is bounded from below) = 1 for each F . Also, for each $F \in \mathfrak{F}$,

$$\begin{aligned} P_F(C[\mu_F]) &= P_F(M - \mu_F \leq Y) \\ &\geq P_F(L_F \leq \alpha S) \\ &\geq P_F(L_F \leq \alpha S, r_F < S) \\ &\geq P_F(L_F \leq \alpha r_F, r_F < S) \\ &\geq P_F(L_F \leq \alpha r_F) + P_F(r_F < S) - 1 \\ &\geq 1 - (1/\alpha) + \beta - 1 \\ &> 0. \end{aligned}$$

This contradiction to Corollary 3 establishes Corollary 4.

The preceding proof consists in showing that if M is an estimator such that $r_F(M)$ is controllable, then μ_F is controllable, contrary to Corollary 3. This argument can also be used to show the uncontrollability of certain parameters. Simple examples of such parameters are the variance of F , the difference between the mean and median values of F , and the supremum of the points of increase of F . Note that while the unboundedness of these parameters is evident when assumptions such as (iii) and (iv) hold, verification that they are uncontrollable is less trivial even in the case when V consists of a single observation.

Finally, let $A(z)$ be a (randomized) function of V taking values in the set of all distribution functions of z . Let $C^*[F]$ denote the event that $A(z) \geq F(z)$ for all z .

THEOREM 2. $\inf_{F \in \mathcal{F}} \{P_F(C^*[F])\} = 0$.

Application of Theorem 2 to $-X_i$ yields with little effort a similar theorem, dual to Theorem 2, concerning the probability that $A(z) \leq F(z)$ for all z . Obviously these two theorems together imply a two-sided version of Theorem 2.

3. Proofs of the theorems. The proofs of the theorems depend on the fact that a given distribution function F can be so modified that, while the probability distribution of the X_i 's (and therewith of V) is perturbed only slightly, parameters such as the mean suffer arbitrary displacements. This modification is described in the following paragraphs, before undertaking the proofs of Theorems 1 and 2.

Let Φ denote the class of all functions φ of V with $0 \leq \varphi \leq 1$, and (for any two distribution functions F and G) define the familiar absolute-variational distance between F and G by

$$(4) \quad \delta(F, G) = \sup_{\varphi \in \Phi} |E_F[\varphi] - E_G[\varphi]|.$$

Given F , let H be an arbitrary distribution function and π an arbitrary constant, $0 < \pi < 1$, and define the distribution function G thus:

$$(5) \quad G(z) = \pi F(z) + (1 - \pi)H(z).$$

The following lemma shows that if the given sampling procedure is closed for F in the sense of (1), and if π is sufficiently close to 1, then, *no matter what H may be*, the probability distributions of V under F and G are not very distant from one another. It may clarify the meaning and proof of the lemma to remark that it is for this application of the lemma, not for the lemma itself, that the sampling procedure must be closed for F .

LEMMA. $\delta(F, G) \leq 1 - \pi^k P_F(N \leq k)$ for each positive integer k .²

PROOF. Choose and fix a positive integer k . Let $R^{(k)}$ denote the space of all points $z^{(k)} = (z_1, z_2, \dots, z_k)$ with $-\infty < z_i < \infty$ for $i = 1, 2, \dots, k$. For any univariate distribution function $F(z)$, write $F^{(k)}(z^{(k)}) = \prod_{i=1}^k F(z_i)$.

It will be shown first that if F and G are related according to (5), then, for any nonnegative function f on $R^{(k)}$,

$$(6) \quad \int_{R^{(k)}} f dG^{(k)} \geq \pi^k \int_{R^{(k)}} f dF^{(k)}.$$

To verify this inequality, let $Y_1, Y_2, \dots, Y_k, Z_1, Z_2, \dots, Z_k$, and U_1, U_2, \dots, U_k be independent random variables such that each Y_i is distributed according to F , each Z_i according to H , and $P(U_i = 1) = 1 - P(U_i = 0) = \pi$

² These inequalities are a considerable improvement of the corresponding ones in an earlier version of the lemma, and the proof is somewhat similar. The authors are indebted to Professor W. Hoeffding for these improvements.

for each U_i . Write $W_i = U_i Y_i + (1 - U_i) Z_i$ for $i = 1, 2, \dots, k$. Then W_1, W_2, \dots, W_k are independent random variables, each distributed according to G as defined by (5). Let B denote the event that $U_i = 1$ for all $i = 1, 2, \dots, k$ and let \bar{B} denote the complement of B . Now it is straightforward to show (6); thus,

$$\begin{aligned}
 \int_{R^{(k)}} f dG^{(k)} &= E[f(W_1, \dots, W_k)] \\
 &= P(B)E[f(W_1, \dots, W_k) | B] + P(\bar{B})E[f(W_1, \dots, W_k) | \bar{B}] \\
 &\geq P(B)E[f(W_1, \dots, W_k) | B] \\
 (7) \quad &= \pi^k E[f(Y_1, \dots, Y_k) | B] \\
 &= \pi^k E[f(Y_1, \dots, Y_k)] \\
 &= \pi^k \int_{R^{(k)}} f dF^{(k)}.
 \end{aligned}$$

Consider the space of all sequences $X^{(\infty)} = (X_1, X_2, \dots \text{ ad inf})$. Since V , the observed sample, is by definition a (randomized) function of $X^{(\infty)}$, it makes sense to speak of the conditional distribution of V given $X^{(k)} = (X_1, \dots, X_k)$. It follows from a well known property of conditional expectation that, for any function $h(V)$ and any F ,

$$(8) \quad E_F[h] = \int_{R^{(k)}} E_F[h | X^{(k)}] dF^{(k)},$$

provided that $E_F[h]$ exists.

Next, let φ be a function of V such that $0 \leq \varphi \leq 1$. Define $\psi(V) = 1$ if $N \leq k$ and $\psi(V) = 0$ if $N > k$. It is easy to see that there exists a function f on $R^{(k)}$ such that $0 \leq f \leq 1$, and

$$(9) \quad E_F[\varphi \cdot \psi | X^{(k)}] = f(X^{(k)}),$$

for all F . The function f depends, of course, on the given φ and the given sampling procedure.

Suppose, now, that F and G are two distribution functions related according to (5). Then,

$$\begin{aligned}
 E_G[\varphi] &\geq E_G[\varphi \cdot \psi] \\
 &= \int_{R^{(k)}} E_G[\varphi \cdot \psi | X^{(k)}] dG^{(k)} && \text{by (8)} \\
 &= \int_{R^{(k)}} f dG^{(k)} && \text{by (9)} \\
 &\geq \pi^k \int_{R^{(k)}} f dF^{(k)} && \text{by (6)}
 \end{aligned}$$

$$= \pi^k \int_{R^{(k)}} E_F[\varphi \cdot \psi | X^{(k)}] dF^{(k)} \quad \text{by (9)}$$

$$= \pi^k E_F[\varphi \cdot \psi] \quad \text{by (8)}$$

$$= E_F[\varphi] - E_F[\varphi(1 - \pi^k \psi)]$$

$$\geq E_F[\varphi] - E_F[1 - \pi^k \psi]$$

$$= E_F[\varphi] - 1 + \pi^k P_F(N \leq k).$$

Thus,

$$(10) \quad E_F[\varphi] - E_G[\varphi] \leq 1 - \pi^k P_F(N \leq k)$$

for all φ in Φ . Since $\varphi \in \Phi$ implies $1 - \varphi \in \Phi$, it follows from (10) that

$$(11) \quad -E_F[\varphi] + E_G[\varphi] \leq 1 - \pi^k P_F(N \leq k)$$

for all φ in Φ . In view of (10), (11), and the definition (4) of δ , $\delta(F, G) \leq 1 - \pi^k P_F(N \leq k)$ for all k , as was to be proved.

PROOF OF THEOREM 1. Let m and m' be real numbers, and let $\epsilon > 0$ be given. Consider a fixed F in \mathfrak{F}_m . Choose and fix a positive integer k such that

$$(12) \quad P_F(N > k) < \epsilon.$$

The existence of such a k is assured by (1). Now choose and fix a π such that $0 < \pi < 1$ and

$$(13) \quad (1 - \pi^k) < \epsilon.$$

Let H be a distribution function in \mathfrak{F} such that $\pi m + (1 - \pi)\mu_H = m'$ (see assumption (ii)), and let G be defined by (5). Then, by assumption (iii), G is in \mathfrak{F} , and since $\mu_G = \pi\mu_F + (1 - \pi)\mu_H = m'$, G is in $\mathfrak{F}_{m'}$. Since $1 - \pi^k P_F(N \leq k) \leq (1 - \pi^k) + P_F(N > k)$, it follows from (12), (13), and the Lemma that $\delta(F, G) < 2\epsilon$.

Since ϵ and F are arbitrary, $\inf_{G \in \mathfrak{F}_{m'}} \{\delta(F, G)\} = 0$ for each $F \in \mathfrak{F}_m$. In other words, $\mathfrak{F}_{m'}$ is everywhere dense in \mathfrak{F}_m , under the metric δ . Since m and m' are arbitrary, it follows (see assumption (i)) that, for each m , \mathfrak{F}_m is everywhere-dense in \mathfrak{F} . This conclusion, together with the observation that $E_F[\varphi]$ is continuous in F for any bounded φ , yields Theorem 1.

PROOF OF THEOREM 2. Before the proof proper we present a line of argument that may be of some interest in suggesting a heuristic connection between this theorem and Theorem 1, though this line of argument makes assumptions that are actually gratuitous. It assumes in fact that (i) obtains and that, for some F , the mean of A almost always exists and is finite.

Suppose, then, that the random distribution function A is such that, for some $F \in \mathfrak{F}$;

$$(14) \quad -\infty < \int_{-\infty}^{\infty} z dA < \infty$$

except for a P_F -null event. Define $I = [\int_{-\infty}^{\infty} z dA, \infty)$ whenever (14) is satisfied, and $I = (-\infty, \infty)$ otherwise. Now, the event $A(z) \geq F(z)$ for all z (that is, $C^*[F]$) implies the event $\infty > \mu_F \geq \int_{-\infty}^{\infty} z dA$ (that is, $C[\mu_F]$) provided only that μ_F exists and is finite. Hence $P_F(C^*[F]) \leq P_F(C[\mu_F])$ for each $F \in \mathfrak{F}$. The desired conclusion now follows from Corollary 3.

Now, dropping the assumptions (i) and (14), turn to the proof proper. Choose and fix an ϵ , $0 < \epsilon < 1$, and an $F \in \mathfrak{F}$ such that $F(0) > 0$. The existence of such an F is assured by assumptions (iv) and (v). For each z , let $J(z) = \inf \{u: P_F(A(z) \leq u) \geq 1 - \epsilon\}$. It is not difficult to see that J is a nondecreasing function of z , with $\lim_{z \rightarrow -\infty} J(z) = 0$, $\lim_{z \rightarrow \infty} J(z) = 1$, and that J is also continuous from the right, so that it is actually a distribution function. Also,

$$(15) \quad P_F\{A(z) > J(z)\} \leq \epsilon$$

for each z .

Now choose k such that (12) holds, choose π such that (13) holds, and choose λ such that $J(\lambda) < (1 - \pi)F(0)$. Let G be defined by (5), with $H(z) \equiv F(z - \lambda)$. Then G is in \mathfrak{F} , by assumptions (iii) and (iv), and

$$(16) \quad J(\lambda) < G(\lambda),$$

by the choice of λ and the definition of G . Hence

$$\begin{aligned} (17) \quad P_G(C^*[G]) &= P_G(A(z) \geq G(z) \text{ for all } z) \\ &\leq P_G(A(\lambda) \geq G(\lambda)) \\ &\leq P_G(A(\lambda) > J(\lambda)) && \text{by (16)} \\ &\leq P_F(A(\lambda) > J(\lambda)) + P_F(N > k) + (1 - \pi^k) \\ &&& \text{by the lemma} \\ &\leq 3\epsilon && \text{by (12), (13), (15).} \end{aligned}$$

Since ϵ is an arbitrary positive fraction, the theorem is proved.

The proof of Theorem 2 does not use quite the full force of (iii) and (iv). It is enough that for some $F \in \mathfrak{F}$ and two sequences of numbers $\alpha_i (0 \leq \alpha_i < 1)$ and β_j , such that $\alpha_i \rightarrow 1$ and $\beta_j \rightarrow \infty$, the distributions G_{ij} such that

$$(18) \quad G_{ij}(z) = \alpha_i F(z) + (1 - \alpha_i)F(z + \beta_j)$$

are in \mathfrak{F} . For the dual of Theorem 2, it is required instead that $\beta_j \rightarrow -\infty$.

REFERENCES

- [1] A. DVORETZKY, A. WALD, AND J. WOLFOWITZ, "Elimination of randomization in certain statistical decision procedures and zero sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1-21.
- [2] ABRAHAM WALD, *Statistical Decision Functions*, New York, John Wiley and Sons, 1950.

ASYMPTOTIC DISTRIBUTIONS OF TWO GOODNESS OF FIT CRITERIA

BY PATRICK BILLINGSLEY

Washington, D. C.

1. Results. Let $\{X_1, X_2, \dots\}$ be a stochastic process in which each random variable takes as values only the integers $1, 2, \dots, s$. To test the null hypothesis that the process is independent and stationary with $P\{X_n = k\} = p_k > 0$, it is natural to form the statistic

$$(1.1) \quad \sum_{u_1, \dots, u_r=1}^s \frac{(n_{u_1 \dots u_r} - np_{u_1} \dots p_{u_r})^2}{np_{u_1} \dots p_{u_r}},$$

where $n_{u_1 \dots u_r}$ is the number of integers $m \leq n$ for which (X_m, \dots, X_{m+r-1}) is the r -tuple (u_1, \dots, u_r) . In Section 2 we show that under the null hypothesis the distribution function of (1.1) approaches, as $n \rightarrow \infty$, the distribution function

$$(1.2) \quad \sum_{\lambda=1}^{r-1} K_{s^{r-1-\lambda}(s-1)^2}(x/\lambda) * K_{s-1}(x/\nu),$$

where $K_i(x)$ is the chi-square distribution with i degrees of freedom and the first $*$ denotes iterated convolution in the obvious way. Good [1], using different methods, has obtained this result for the special case in which the p_k are all equal and s is a prime number.

If the p_k are estimated by n_k/n , there results the statistic

$$(1.3) \quad \sum_{u_1, \dots, u_r=1}^s \frac{(n_{u_1 \dots u_r} - n^{1-r} n_{u_1} \dots n_{u_r})^2}{n^{1-r} n_{u_1} \dots n_{u_r}}.$$

In Section 3 we show that under the hypothesis that $\{X_n\}$ is stationary and independent, the distribution function of (1.3) approaches, as $n \rightarrow \infty$, the distribution function

$$(1.4) \quad \sum_{\lambda=1}^{r-1} K_{s^{r-1-\lambda}(s-1)^2}(x/\lambda).$$

In the special case $\nu = 2$ this result is implicit in the work of Hoel [2]. Note that in this case (1.4) becomes $K_{(s-1)^2}(x)$.

The means and variances of the distributions (1.2) and (1.4) are easily written down. It is obvious that if ν is fixed and $s \rightarrow \infty$, then these distributions are, when normed by their means and standard deviations, asymptotically normal. It is a simple matter to show, using Ljapunov's condition and the fact that the distributions are convolutions, that the same thing is true if s is fixed and $\nu \rightarrow \infty$. By interpolation in the tables of [3] one can get an approximation to (1.2) for the case $s = 2$ and $\nu = 2$ and an approximation to (1.4) for the case

Received November 22, 1955.

$s = 2$ and $\nu = 3$. The paper [3] also deals with the general problem of computing and approximating the distributions of weighted sums of independent chi-square-distributed random variables.

The author wishes to thank E. H. Spanier for some essential ideas on the algebraic aspects of this work.

In his paper "On the serial test for random sequences," forthcoming in this journal, I. J. Good shows that the expected value of the statistic (1.1) and the first moment of its limiting distribution (1.2) have the same value, viz., $s^r - 1$.

2. Asymptotic distribution of (1.1). In what follows we make use of the theory of finite dimensional vector spaces. The reader is referred to [4] for the notions of direct sum (denoted here by \vee), spectral decomposition, etc. We denote an operator and the matrix which represents it by the same symbol.

Let \mathcal{U}_r be s^r -dimensional Euclidean space with components indexed by the s^r ν -tuples (u_1, \dots, u_r) with $1 \leq u_i \leq s$. Let

$$(2.1) \quad \rho_{u_1 \dots u_r} = (p_{u_1} \dots p_{u_r})^{1/2},$$

and let x be the random vector in \mathcal{U}_r with components

$$x_{u_1 \dots u_r} = (n_{u_1 \dots u_r} - n \rho_{u_1 \dots u_r}^2) / n^{1/2} \rho_{u_1 \dots u_r}.$$

Then $|x|^2$ is the statistic (1.1). Let $M^{(r)}$ be the s^r by s^r matrix with entries defined by

$$M_{u_1 \dots u_r, v_1 \dots v_r}^{(r)} = \delta_{u_1, v_1} \dots \delta_{u_r, v_r} - \rho_{u_1 \dots u_r} \rho_{v_1 \dots v_r}.$$

For this matrix to be well defined, the ν -tuples must be ordered. Which order is taken is immaterial so long as it is kept constant throughout the argument.

We show first of all that the covariance matrix of x is asymptotically $\Lambda^{(r)}$, where $\Lambda^{(r)}$ is the s^r by s^r matrix with entries defined by

$$(2.2) \quad \Lambda_{u_1 \dots u_r, v_1 \dots v_r}^{(r)} = M_{u_1 \dots u_r, v_1 \dots v_r}^{(r)} + \sum_{k=1}^{r-1} \rho_{u_1 \dots u_k} \rho_{v_1 \dots v_k} M_{u_{k+1} \dots u_r, v_{k+1} \dots v_r}^{(r-k)} \\ + \sum_{k=1}^{r-1} \rho_{u_1 \dots u_{r-k+1}} \rho_{v_1 \dots v_{r-k+1}} M_{u_1 \dots u_{r-k}, v_1 \dots v_{r-k}}^{(r-k)}.$$

Let $\alpha_i[\beta_i]$ be 1 or 0 according as $(X_i, \dots, X_{i+\nu-1})$ is the ν -tuple (u_1, \dots, u_r) $[(v_1, \dots, v_r)]$ or not. Then $n_{u_1 \dots u_r} = \sum_{i=1}^n \alpha_i$ and $n_{v_1 \dots v_r} = \sum_{i=1}^n \beta_i$. Let $c(i, j) = \text{cov}(\alpha_i, \beta_j)$. Then $c(i, j) = 0$ if $|i - j| \geq \nu$ and $c(i, i + k)$ is independent of i . Hence, since $|c(i, j)| \leq 2$, we have

$$\begin{aligned} \text{cov}(n_{u_1 \dots u_r}, n_{v_1 \dots v_r}) &= \sum_{i=1}^n \sum_{j=1}^n c(i, j) \\ &= \sum_{i=1}^n c(i, i) + \sum_{i=1}^n \sum_{k=1}^{r-1} (c(i, i + k) + c(i + k, i)) \\ &\quad - \sum_{i=n-\nu+2}^n \sum_{k=n-i+1}^{r-1} (c(i, i + k) + c(i + k, i)) \\ &= n[c(1, 1) + \sum_{k=1}^{r-1} (c(1, 1 + k) + c(1 + k, 1))] + 2\theta\nu^2 \end{aligned}$$

with $|\theta| \leq 1$. Hence in the limit $\text{cov}(n_{u_1 \dots u_r}, n_{v_1 \dots v_r})$ is

$$\lim_{n \rightarrow \infty} \Lambda_{u_1 \dots u_r, v_1 \dots v_r}^{(\nu)} = \rho_{u_1 \dots u_r}^{-1} \rho_{v_1 \dots v_r}^{-1} [c(1, 1) + \sum_{k=1}^{r-1} (c(1, 1+k) + c(1+k, 1))].$$

But for $k = 1, \dots, r-1$,

$$c(1, 1+k) = \delta_{u_{k+1}, v_1} \dots \delta_{u_r, v_{r-k}} p_{u_1} \dots p_{u_r} p_{v_{r-k+1}} \dots p_{v_r} - \rho_{u_1 \dots u_r}^2 \rho_{v_1 \dots v_r}^2.$$

From this expression and similar ones for $c(1, 1)$ and $c(1+k, 1)$, (2.2) follows.

It is an immediate consequence of the multivariate central limit theorem for ν -dependent random variables [5] that the distribution of x approaches that normal distribution having zero means and having $\Lambda^{(\nu)}$ as covariance matrix.

We proceed now to find the spectral decomposition of $\Lambda^{(\nu)}$. For $\nu > 1$ let \mathcal{L}_r be the set of $t \in \mathcal{U}_r$ satisfying

$$(2.3) \quad \sum_{u_1 \dots u_r} \rho_{u_1 \dots u_r} t_{u_1 \dots u_r} = 0$$

and

$$(2.4) \quad \sum_{u_1} \rho_{u_1} t_{u_1 \dots u_r} = \sum_{u_1} \rho_{u_1} t_{u_2 \dots u_r, u_1}.$$

for all (u_2, \dots, u_r) . It follows from (2.3) and (2.4) that for $k = 1, \dots, r$ and $i = 1, \dots, k$,

$$(2.5) \quad \sum_{u_1 \dots u_k} \rho_{u_1 \dots u_k} t_{u_1 \dots u_r} = \sum_{u_1 \dots u_k} \rho_{u_1 \dots u_k} t_{u_{i+1} \dots u_r, u_1 \dots u_i}.$$

Let \mathcal{L}_1 be the set of $t \in \mathcal{U}_1$ for which $\sum_u \rho_u t_u = 0$ and let \mathcal{L}_0 consist of the number 0 alone. For $\nu \geq 1$, define a linear mapping $\Pi_r: \mathcal{L}_r \rightarrow \mathcal{L}_{r-1}$ by $(\Pi_r t)_{u_1 \dots u_{r-1}} = \sum_u \rho_u t_{u u_1 \dots u_{r-1}}$. That $t \in \mathcal{L}_r$ implies $\Pi_r t \in \mathcal{L}_{r-1}$ can be verified by computation. For $\nu \geq 2$, define a second linear mapping $\Omega_{r-1}: \mathcal{L}_{r-1} \rightarrow \mathcal{L}_r$ by

$$(2.6) \quad (\Omega_{r-1} t)_{u_1 \dots u_r} = \rho_{u_1} t_{u_2 \dots u_r} + \rho_{u_r} t_{u_1 \dots u_{r-1}} - \rho_{u_1 u_r} (\Pi_{r-1} t)_{u_2 \dots u_{r-1}}.$$

If $\nu = 2$ the last term in (2.6) is to be omitted. Again a computation shows that $\Omega_{r-1} t \in \mathcal{L}_r$, if $t \in \mathcal{L}_{r-1}$. From these definitions it follows that

$$(2.7) \quad \Pi_r \Omega_{r-1} t = t, \quad t \in \mathcal{L}_{r-1}.$$

Let \mathcal{L}_r^0 be the set of $t \in \mathcal{L}_r$ such that $\Pi_r t = 0$. Then

$$(2.8) \quad \mathcal{L}_r = \mathcal{L}_r^0 \vee \Omega_{r-1}(\mathcal{L}_{r-1}).$$

In fact if $t \in \mathcal{L}_r$, then $t = (t - \Omega_{r-1} \Pi_r t) + \Omega_{r-1} \Pi_r t$, while $t - \Omega_{r-1} \Pi_r t \in \mathcal{L}_r^0$ and $\Omega_{r-1} \Pi_r t \in \Omega_{r-1}(\mathcal{L}_{r-1})$. And if $t \in \mathcal{L}_r^0 \cap \Omega_{r-1}(\mathcal{L}_{r-1})$, then $t = \Omega_{r-1} t'$ and $0 = \Pi_r \Omega_{r-1} t' = t'$ so $t = 0$.

If $t \in \mathcal{L}_\nu$ then $M^{(\nu)}t = t$. From this and (2.5) it follows that

$$\begin{aligned} \sum_{v_1, \dots, v_\nu} \rho_{v_{\nu-k+1}, \dots, v_\nu} M^{(\nu-k)}_{u_{k+1}, \dots, u_\nu, v_1, \dots, v_{\nu-k}} t_{v_1, \dots, v_\nu} \\ = \sum_{v_1, \dots, v_{\nu-k}} M^{(\nu-k)}_{u_{k+1}, \dots, u_\nu, v_1, \dots, v_{\nu-k}} (\Pi_{\nu-k+1} \cdots \Pi_\nu t)_{v_1, \dots, v_{\nu-k}} = (\Pi_{\nu-k+1} \cdots \Pi_\nu t)_{u_{k+1}, \dots, u_\nu}. \end{aligned}$$

Using this relation and the symmetric one, we get

$$\begin{aligned} (\Lambda^{(\nu)} t)_{u_1, \dots, u_\nu} &= t_{u_1, \dots, u_\nu} + \sum_{k=1}^{\nu-1} \rho_{u_1, \dots, u_k} (\Pi_{\nu-k+1} \cdots \Pi_\nu t)_{u_{k+1}, \dots, u_\nu} \\ (2.9) \quad &+ \sum_{k=1}^{\nu-1} \rho_{u_{\nu-k+1}, \dots, u_\nu} (\Pi_{\nu-k+1} \cdots \Pi_\nu t)_{u_1, \dots, u_{\nu-k}} \end{aligned}$$

for $t \in \mathcal{L}_\nu$. If $\nu \geq 3$ and $t \in \mathcal{L}_{\nu-1}$, then by (2.7) and (2.9) we have

$$\begin{aligned} ((\Lambda^{(\nu)} \Omega_{\nu-1} - \Omega_{\nu-1}) t)_{u_1, \dots, u_\nu} &= \rho_{u_1} t_{u_2, \dots, u_\nu} + \sum_{k=2}^{\nu-1} \rho_{u_1, \dots, u_k} (\Pi_{\nu-k+1} \cdots \Pi_{\nu-1} t)_{u_{k+1}, \dots, u_\nu} \\ &+ \rho_{u_\nu} t_{u_1, \dots, u_{\nu-1}} + \sum_{k=2}^{\nu-1} \rho_{u_{\nu-k+1}, \dots, u_\nu} (\Pi_{\nu-k+1} \cdots \Pi_{\nu-1} t)_{u_1, \dots, u_{\nu-k}}. \end{aligned}$$

From this, using (2.9) again, one shows by a long but straightforward calculation that for $\nu \geq 3$

$$(2.10) \quad \Lambda^{(\nu)} \Omega_{\nu-1} - \Omega_{\nu-1} = \Omega_{\nu-1} \Lambda^{(\nu-1)}$$

on $\mathcal{L}_{\nu-1}$.

We next show that for $\nu \geq 2$

$$(2.11) \quad \Lambda^{(\nu)} = I + \sum_{k=2}^{\nu} \Omega_{\nu-1} \cdots \Omega_{k-1} \Pi_k \cdots \Pi_\nu,$$

on \mathcal{L}_ν . The proof goes by induction. The verification being simple for $\nu = 2$, assume (2.11) holds with ν replaced by $\nu - 1$. Then by (2.10) and (2.7) we have

$$\Lambda^{(\nu)} \Omega_{\nu-1} = \left(I + \sum_{k=2}^{\nu} \Omega_{\nu-1} \cdots \Omega_{k-1} \Pi_k \cdots \Pi_\nu \right) \Omega_{\nu-1}.$$

In other words, it follows that (2.11) holds on $\Omega_{\nu-1}(\mathcal{L}_{\nu-1})$. Since it obviously holds on \mathcal{L}_ν^0 , it follows by (2.8) that (2.11) holds on all of \mathcal{L}_ν .

Let $\mathfrak{M}_1 = \mathcal{L}_\nu^0$ and for $\lambda = 2, \dots, \nu$ let

$$(2.12) \quad \mathfrak{M}_\lambda = \Omega_{\nu-1} \cdots \Omega_{\nu+1-\lambda} \mathcal{L}_{\nu+1-\lambda}^0.$$

It follows from (2.8) by induction that $\mathcal{L}_\nu = \mathfrak{M}_1 \vee \cdots \vee \mathfrak{M}_\nu$. Using (2.11) one easily shows that for $\lambda = 1, \dots, \nu$,

$$(2.13) \quad \Lambda^{(\nu)} t = \lambda t \quad \text{if} \quad t \in \mathfrak{M}_\lambda.$$

Let $\sigma \in \mathcal{U}_\nu$ be the vector whose (u_1, \dots, u_ν) -th component is ρ_{u_1, \dots, u_ν} and let $\sigma(v_1, \dots, v_{\nu-1}) \in \mathcal{U}_\nu$ be the vector whose (u_1, \dots, u_ν) -th component is

$\rho_{u_1} \delta_{u_2, v_1} \cdots \delta_{u_r, v_{r-1}} - \rho_{u_r} \delta_{u_1, v_1} \cdots \delta_{u_{r-1}, v_{r-1}}$. Let \mathfrak{M}_0 be the manifold generated by σ and the s^{r-1} vectors $\sigma(v_1, \dots, v_{r-1})$. By definition \mathcal{L}_r is the orthogonal complement of \mathfrak{M}_0 , so that

$$(2.14) \quad \mathfrak{U}_r = \mathfrak{M}_0 \vee \mathfrak{M}_1 \vee \cdots \vee \mathfrak{M}_r.$$

Direct computations show that $\Lambda^{(s)} \sigma = \Lambda^{(s)} \sigma(v_1, \dots, v_{r-1}) = 0$, so that (2.13) holds for $\lambda = 0$. Thus each \mathfrak{M}_λ , $0 \leq \lambda \leq r$, consists of eigenvectors with eigenvalue λ . These are all the invariant subspaces, in view of (2.14).

We now compute the dimension of \mathcal{L}_r^0 . It is easy to show that $\dim \mathcal{L}_1^0 = s - 1$. Suppose $r \geq 2$. To say that $t \in \mathcal{L}_r^0$ is to say that for all u_1, \dots, u_{r-1}

$$(2.15) \quad \sum_u \rho_u t_{uu_1 \cdots u_{r-1}} = \sum_u \rho_u t_{u_1 \cdots u_{r-1} u} = 0.$$

Let X be the s^{r-1} by s^r matrix with entries

$$X_{u_1 \cdots u_{r-1}, v_1 \cdots v_r} = \rho_{v_1} \delta_{u_1, v_2} \cdots \delta_{u_{r-1}, v_r}$$

and let Y be the s^{r-1} by s^r matrix with entries

$$Y_{u_1 \cdots u_{r-1}, v_1 \cdots v_r} = \rho_{v_r} \delta_{u_1, v_1} \cdots \delta_{u_{r-1}, v_{r-1}}.$$

The partitioned matrix

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}$$

is the matrix of the system (2.15), i.e., t lies in \mathcal{L}_r^0 if and only if t is orthogonal to each row of Z . In order to find the (column) rank of Z let A and B be column vectors with s^{r-1} components $A_{u_1 \cdots u_{r-1}}$ and $B_{u_1 \cdots u_{r-1}}$ respectively, and let C be the partitioned vector

$$C = \begin{bmatrix} A \\ B \end{bmatrix}.$$

Now C is orthogonal to each column of Z if and only if $\rho_{u_1} A_{u_2 \cdots u_r} = -\rho_{u_r} B_{u_1 \cdots u_{r-1}}$. Thus if for a set $\{D_{u_2 \cdots u_{r-1}}\}$ of s^{r-2} numbers we let $A_{u_2 \cdots u_r} = \rho_{u_r} D_{u_2 \cdots u_{r-1}}$ and $B_{u_1 \cdots u_{r-1}} = -\rho_{u_1} D_{u_2 \cdots u_{r-1}}$, then C is orthogonal to the columns of Z . Conversely, if C is orthogonal to these columns, it can be cast in this form. Hence the subspace of $2s^{r-1}$ -dimensional space orthogonal to the subspace generated by the columns of Z has dimension s^{r-2} . Therefore Z has rank $2s^{r-1} - s^{r-2}$ and $\dim \mathcal{L}_r^0 = s^{r-2}(s-1)^2$. It now follows by (2.12) and (2.14) that $\dim \mathfrak{M}_0 = s^{r-1}$, $\dim \mathfrak{M}_r = s - 1$, and $\dim \mathfrak{M}_\lambda = s^{r-1-\lambda}(s-1)^2$ for $\lambda = 1, 2, \dots, r-1$.

We now have the dimensions of the invariant subspaces and hence the multiplicities of the eigenvalues of $\Lambda^{(s)}$. Since the distribution of x is asymptotically normal with covariance matrix, $\Lambda^{(s)}$, it follows by an obvious generalization of the result of Section 24.5 of [6] that the distribution of $|x|^2$, or (1.1), approaches (1.2), under the null hypothesis.

3. Asymptotic distribution of (1.3). We assume now that $\{X_n\}$ is independent and stationary, but we regard the p_k as unknown. In fact, let p_1, \dots, p_{s-1} be parameters to be estimated, define p_s by $p_s = 1 - \sum_{k=1}^{s-1} p_k$, and let ρ_{u_1, \dots, u_s}^2 be a function of p_1, \dots, p_{s-1} defined by (2.1).

Now it is easy to show that the values p_k which maximize

$$\prod_{u_1, \dots, u_s} (\rho_{u_1, \dots, u_s}^2)^{n_{u_1, \dots, u_s}}$$

are $p_k = n_k/n + \epsilon_k$, where $\epsilon_k = O(1/n)$. And now the reasoning of Section 30.3 of [6] becomes applicable. Let B be the s' by $s-1$ matrix with entries $B_{u_1, \dots, u_s, u} = \rho_{u_1, \dots, u_s}^{-1} \partial \rho_{u_1, \dots, u_s}^2 / \partial p_u$, $1 \leq u_i \leq s$, $1 \leq u < s$. Let y be the random vector which results from substituting the estimate n_k/n for p_k in x . Then $|y|^2$ is (1.3). In order that the theorem of Section 30.3 of [6] be directly applicable it would be necessary that $\{n_{u_1, \dots, u_s}\}$ be a sample from a multinomial universe. However, since the x defined above is asymptotically normal with covariance matrix $\Lambda^{(v)}$, since B has rank $s-1$ and since $x_{u_1, \dots, u_s} = o(n^{1/4})$ in probability (as is easily shown), a perusal of the proof of the theorem referred to shows that we are in the present case justified in concluding that the distribution of y approaches that normal distribution with zero means and covariance matrix $A\Lambda^{(v)}A'$, where $A = I - B(B'B)^{-1}B'$.

We now find the spectral decomposition of $A\Lambda^{(v)}A'$. Let K be the s by $s-1$ matrix with entries $K_{u,v}$, where $K_{u,v} = \delta_{u,v}$ if $u < s$ and $K_{s,v} = -1$. Let J be the s' by s matrix with entries

$$J_{u_1, \dots, u_s, u} = \sum_{i=1}^s \rho_{u_1, \dots, u_{i-1} u_{i+1}, \dots, u_s} \delta_{u_i, u} \rho_{u_i, u}^{-1}.$$

Then $B = JK$.

If $t \in \mathcal{L}_s$ it follows from (2.5) that $(J't)_u = \nu \rho_u^{-1} (\Pi_2 \cdots \Pi_s t)_u$. From this it follows that $J't = 0$ for $t \in \mathfrak{M}_1 \vee \cdots \vee \mathfrak{M}_{s-1}$. If $\sigma(v_1, \dots, v_{s-1})$ is defined as in Section 2, then, as a direct computation shows, $J'\sigma(v_1, \dots, v_{s-1}) = 0$. Moreover, $(J'\sigma)_u = \nu$, so that $B'_\sigma = 0$. Hence $B't = 0$ for $t \in \mathfrak{M}_0 \vee \cdots \vee \mathfrak{M}_{s-1}$. Now the matrix A is symmetric and idempotent, so that, viewed as an operator, it is a perpendicular projection on the manifold $\mathfrak{R} = \{t: At = t\} = \{t: B(B'B)^{-1}B't = 0\}$. It is easy to show that the rank of $B(B'B)^{-1}B'$ is the same as that of B , viz., $s-1$. Hence $\dim \mathfrak{R} = s' - s + 1$. We have shown that $\mathfrak{M}_0 \vee \cdots \vee \mathfrak{M}_{s-1} \subset \mathfrak{R}$, and since $\dim (\mathfrak{M}_0 \vee \cdots \vee \mathfrak{M}_{s-1}) = s' - s + 1$ (cf. Section 2), we have $\mathfrak{M}_0 \vee \cdots \vee \mathfrak{M}_{s-1} = \mathfrak{R}$. The manifolds \mathfrak{M}_λ , being the invariant spaces of the symmetric matrix $\Lambda^{(v)}$, are mutually orthogonal. Hence \mathfrak{M}_s is the orthogonal complement of \mathfrak{R} and $At = 0$ for $t \in \mathfrak{M}_s$. Therefore $A\Lambda^{(v)}A't = \lambda t$ if $t \in \mathfrak{M}_\lambda$ with $1 \leq \lambda < s$, while $A\Lambda^{(v)}A't = 0$ if $t \in \mathfrak{M}_0 \vee \mathfrak{M}_s$. Finally $\dim \mathfrak{M}_\lambda = s^{s-1-\lambda}(s-1)^2$ for $\lambda = 1, \dots, s-1$ and $\dim \mathfrak{M}_0 \vee \mathfrak{M}_s = s^{s-1} + s - 1$.

Thus we have the eigenvalues of $A\Lambda^{(v)}A'$, with their multiplicities, and it follows as in Section 2 that the distribution of $|y|^2$, or (1.3), approaches (1.4).

REFERENCES

- [1] I. J. GOOD, "The serial test for random sampling numbers and other tests for randomness," *Proc. Camb. Phil. Soc.*, 49₂ (1953), pp. 276-284.
- [2] P. G. HOEL, "A test for Markov chains," *Biometrika*, 41 (1954), pp. 430-433.
- [3] A. GRAD AND H. SOLOMON, "Distribution of quadratic forms and some applications," *Ann. Math. Stat.*, 26 (1955), pp. 464-477.
- [4] P. R. HALMOS, *Finite Dimensional Vector Spaces*, *Annals of Math. Studies*, No. 7.
- [5] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, 15 (1948), pp. 773-780.
- [6] H. Cramér, *Mathematical Methods of Statistics*, Princeton, 1946.

THE SURPRISE INDEX FOR THE MULTIVARIATE NORMAL DISTRIBUTION

By I. J. Good

1. The surprise index and its generalisations. Let E_1, E_2, E_3, \dots be a natural classification into a finite or countably infinite number of possible mutually exclusive and exhaustive results of some experiment or observation, and let $P(E_i | H) = p_i (i = 1, 2, 3, \dots)$, where H is a simple statistical hypothesis. Then the surprise index (Weaver [7]) associated with the result E_i is

$$(1) \quad \lambda_i = \frac{E(p_i | H)}{p_i} = \frac{\sum_j p_j^2}{p_i}.$$

If the experiment consists in the measurement of a continuous vector or scalar variable with a differentiable distribution function, we define

$$(2) \quad \lambda_i = \frac{E(p^* | H)}{p},$$

where p^* is the random variable that is the probability density of the original random variable, and p is a realisation of p^* .

For practical purposes, (2) is almost the same definition as (1). For example, a continuous scalar variable is usually measured to some fixed number, n , of decimal places, and the natural classification of the possible results of the experiment is into intervals of length 10^{-n} of values of the variable. If we then use definition (1) and let n tend to infinity, we get definition (2). For experiments with results that are real variables having distributions that are partly discrete (atomic) and partly continuous (differentiable), it is not immediately obvious what definition should be used. Something more will be said about this later.

The surprise index is open to two criticisms:

(I) It is changed when the results of an experiment are lumped together in a new way, in the discrete case, or when there is a change of mathematically independent variable in the continuous case.

(II) The numerator in (1) or (2) is somewhat arbitrary.

We shall now discuss these two criticisms.

As an example of (I), suppose that an "unbiased coin" is spun twenty times.¹ There is an obvious classification of the possible results of the experiment into 2^{20} categories. But, with this classification,

$$(3) \quad H T H T H T H T H T H T H T H T H T H T$$

Received September 22, 1955.

¹ By putting the description "unbiased coin" in quotation marks, we intend to imply that a certain self-explanatory simple statistical hypothesis is to be taken for granted, and the probabilities of the possible results are the tautological ones usually associated with this idealised experiment.

has the same surprise index as

(4) *HTTHTHHHTTHHHHTTTHTH.*

In practice, (3) would be more surprising than (4), at any rate if neither of them had been written down in advance of the experiment. This is partly because (3) is simpler. (The reader should avoid being confused by the two meanings of the word "simple." We use the word in its technical sense only in the phrase "simple statistical hypothesis," while in "simple hypothesis," the word has its ordinary non-technical meaning.)

If we imagine that the 2^{20} possible results are classified into groups of roughly equal simplicity, (3) would belong to a small group, whereas (4) would belong to a large group. If we regard all the results in one group as a single possible result, it follows that (3) would, after all, have a higher surprise index than (4). Thus the vagueness of definition (1) is seen to arise from the difficulty of measuring simplicity. (With regard to the regrouping of results, see Bartlett [1], page 231.)

The connection between surprise and simplicity can be defended by the following argument.

Perhaps the main biological function of surprise is to jar us into reconsidering the validity of some hypothesis that we had previously accepted. Hence, we tend to be surprised when evidence is received against such a hypothesis, i.e., when the result of an observation has much greater probability when given some other, not entirely untenable, hypothesis. But in the process of being surprised, we often do not have time to estimate the initial probability of the rival hypothesis; instead, we tend to notice whether the rival hypothesis is very simple. More formally, we are surprised if E occurs when the likelihood ratio $P(E | H') / P(E | H)$ is large, where H was previously believed and H' is very simple.

Fortunately, simple hypotheses often have higher initial credibilities than complicated ones, so that the capacity of surprise leads to the discovery of new truths.

In the above example, a hypothesis that would explain (3) would be that the coin always, or very often, rotates by the same (odd) number of half-revolutions.

Since no one has yet thought of a satisfactory measure of simplicity, it seems unlikely that a really satisfactory measure of surprise can be given. For an experiment whose result is naturally expressed as a single integer, the difficulty does not seem to matter greatly. It is true that we may be temporarily surprised because the integer has striking properties, like those of 10,000 or 22,222, but we are often able to discount this sort of surprise as being due to a "mere coincidence" and as being dependent on the irrelevancy that we use radix 10.

Obvious examples of experiments whose results are integers are those giving rise to binomial and Poisson distributions. For these, λ_1 , has been evaluated by Redheffer [6]: for the binomial distribution, λ_1 is expressible in terms of the

sums of the squares of the binomial terms (not coefficients) and therefore in terms of Legendre polynomials. Outside the range of existing tables, the Legendre polynomials that occur here may be conveniently computed with the help of a formula given by Good [4].

We now consider criticism (II). A generalisation of the surprise index, with a more general numerator, has been briefly discussed by Good [3]. Let

$$\lambda_0 = \frac{[E(p^*)]^{1/u}}{p} \quad (u > 0),$$

$$\lambda_0 = \exp \{E(\log p^*) - \log p\} = \text{G.E.}(p^*)/p,$$

(where G.E. means "geometric expectation"), and let

$$\Lambda_u = \log \lambda_u \quad (u \geq 0).$$

We may call Λ_u a "logarithmic surprise index." It can be seen at once that $\lambda_u (u \geq 0)$ is multiplicative, whereas Λ_u is additive, if the results of several statistically independent experiments are combined into a single experiment. Weaver [7] did not allow his surprise index to be less than 1, but it is necessary to do so in order to achieve multiplicativity. A negative logarithmic surprise index corresponds to an event that "was only to be expected."

Of the continuous infinity of surprise indexes, the most natural ones seem to be λ_1 and λ_0 , or, equivalently, Λ_1 and Λ_0 . Bartlett [1] discussed Λ_0 , but not in relation to Weaver's suggestion. We shall argue below that λ_0 (or Λ_0) is rather better than λ_1 , at any rate for multivariate normal distributions. For univariate normal distributions, there is little difference between Λ_0 and Λ_1 .

Before going on to this, we shall digress for a moment in order to discuss (i) distributions that are partly discrete (as promised earlier) and (ii) tail-area probabilities.

2. Partly discrete distributions. The above reference to multiplicativity suggests a possible definition of λ_u for a univariate distribution that is partly discrete and partly continuous. We can first classify the possible results into "atomic" on the one hand and "non-atomic" on the other. This is a two-category (discrete) classification for which $\lambda_u^{(1)}$ may be defined as λ_u was before. Then, if the observed value of the random variable is atomic (or non-atomic), we can compute a conditional $\lambda_u^{(2)}$, i.e., conditional on the information that the variable is an atomic one (or a non-atomic one). Finally, we can define $\lambda_u = \lambda_u^{(1)} \lambda_u^{(2)}$.

3. Tail-area probabilities. The so-much-or-more method in statistics is the usual method in which the result of an experiment or observation is summarised by means of a tail-area probability

$$P(x^* > x), \quad P(x^* \geq x), \quad \text{or} \quad P(x^* > x) + \frac{1}{2}P(x^* = x),$$

where x^* is a real random variable and x is a real number. This method is most satisfying when x^* is the likelihood of an experiment (given a null hypothesis), but in this case the tail-area probability is often difficult to evaluate numerically.

Moreover, there are again logical difficulties for distributions that are partly discrete and partly continuous.

The reciprocal of a tail-area probability is often not more than about 10 times the Bayes factor against the null hypothesis, calculated in accordance with some reasonable assumptions about the initial distributions and probabilities (See Good [2], page 94.) When the ratio is greater than about 10, there is likely to be some argument about which is the better statistic. This difficulty can easily arise for bimodal distributions.

Jeffreys [5], page 316, says "What the use of P (a tail-area probability) implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred." In other words, a tail-area probability consists in the probability of an experimental result artificially added to the probabilities of results that did not occur, or, if not artificially, at any rate with incomplete logical justification.

λ_u , for any u , as a final summary of an experiment or observation, overcomes Jeffreys' criticism of a tail-area probability, although it may still be unsatisfactory as compared with upper and lower bounds for a Bayes factor when we are prepared to assume enough about the non-null hypothesis. For a distribution with density such as

$$\frac{1}{\sqrt{8\pi}} (e^{-(x+4)^2/2} + e^{-(x-4)^2/2}),$$

λ_u is apt to be a much better summary of the experiment than a tail-area probability would be. But it can be argued that better still would be the tail-area probability associated with the value of λ_u . This would come to the same thing as the use of the distribution of the likelihood or the likelihood density. (The possibility of using Weaver's surprise index, λ , as a substitute for the use of tail-area probabilities was suggested in conversation by Mr. G. C. Wall.)

4. λ_u for multivariate normal distributions. For multivariate normal distributions, $P(p^* < p)$, the distribution of the likelihood density, does not seem to be expressible in elementary terms. It is therefore perhaps more worth while to compute λ_u for the multivariate normal distribution than for the Poisson and binomial distributions.

A k -dimensional multivariate normal distribution has a density function of the form

$$p = \frac{|A|^{1/2}}{(2\pi)^{1/2k}} \exp \left\{ -\frac{1}{2} \sum_{i,j}^{1,2,\dots,k} A_{ij}(x_i - a_i)(x_j - a_j) \right\},$$

where $|A| = \det\{A_{ij}\}$. (See, for example, Wilks [8], p. 65.)

Now, it is easily seen that for any k -dimensional probability density, the generalised surprise indexes λ_u , $\Lambda_u(u \geq 0)$ are invariant under all non-singular linear transformations.² This observation follows from the fact that the Jacobian

² The method of this paragraph is due to the referee; my own method was clumsier.

of such a transformation is constant and non-zero. Therefore, for non-degenerate k -dimensional multivariate normal distributions, there is no real loss of generality in taking $A = I$ (the identity matrix) and $a_1 = a_2 = \dots = a_k = 0$. For this standardised distribution, we can use the multiplicative property of λ_u for probabilistically independent experiments, together with a simple univariate integration, to evaluate λ_u . Then, transforming back to the general non-singular distribution, we get

$$\lambda_u = \frac{1}{(u+1)^{k/2u}} \exp \left\{ \frac{1}{2} \sum_{i,j} A_{ij}(x_i - a_i)(x_j - a_j) \right\} \quad (u > 0),$$

$$\Lambda_u = \frac{1}{2} \sum_{i,j} A_{ij}(x_i - a_i)(x_j - a_j) - \frac{k}{2u} \log(u+1) \quad (u > 0),$$

$$\Lambda_0 = \frac{1}{2} \left\{ \sum_{i,j} A_{ij}(x_i - a_i)(x_j - a_j) - k \right\}.$$

It may be observed that Λ_u (and therefore λ_u), regarded as a function of u , is continuous to the right at $u = 0$. By writing $u = e^v - 1$, we see at once that Λ_u is a strictly increasing function of u . When $u \rightarrow \infty$, Λ_u tends to

$$\Lambda_\infty = \frac{1}{2} \sum_{i,j} A_{ij}(x_i - a_i)(x_j - a_j).$$

From this expression it is clear that Λ_∞ is the logarithm of the likelihood ratio in the sense of Wilks [8] for testing the hypothesis of our multivariate normal distribution "within" the more general class of multivariate normal distributions that have the same matrix $\{A_{ij}\}$, or, what comes to the same thing, the same covariance matrix.

It is known (see, for example, Wilks [8], page 104) that $2\Lambda_\infty$ has precisely a chi-squared (gamma-variate) distribution with k degrees of freedom. Since

$$\Lambda_u = \Lambda_\infty - \frac{k}{2u} \log(u+1),$$

we can obtain the exact tail-area probability corresponding to any observed value of Λ_u . But we may also develop an intuitive appreciation of Λ_u (or λ_u) in itself, for some fixed value of u . In order to decide which is the most natural value of u to take, we note that $E(\Lambda_0) = 0$. (This is obvious from the definition of Λ_0 and also from the fact that $2\Lambda_0 + k$ has a chi-squared distribution with k degrees of freedom.) It seems natural to demand that the expected log-surprise should be zero before an experiment is performed. It is not equally natural to insist that $E(\lambda_u) = 1$, or that $E(\lambda_u^{-1}) = 1$ (which gives $u = 1$), since λ_u and λ_u^{-1} have very skew distributions. For very skew distributions, expected values are more artificial than for ordinary distributions such as the chi-squared. For one thing, the median is a long way from the expected value for very skew distributions.

We conclude, then, that for the k -dimensional multivariate normal distribution,

Λ_0 and λ_0 seem more natural measures of surprise than Λ_1 and λ_1 , whereas other values of u do not seem to have anything special to commend them. There is little difference between Λ_0 and Λ_1 when k is small.

For $k = 1$, we have

$$\lambda_0 = e^{s^2/2} / \sqrt{e}, \quad \lambda_1 = e^{s^2/2} / \sqrt{2},$$

where s is the "sigma-age" of an observation; i.e., the deviation from the mean divided by the standard deviation. Some numerical values of λ_0 and λ_1 are given in the following table, together with the reciprocals of the corresponding two-tailed tail-area probabilities, $P(s)$.

s	0	1	2	3	4	5
$1/P(s)$	1	3.1	22	370	16000	1740000
λ_0	0.61	1	4.5	54.6	1800	160000
λ_1	0.71	1.17	5.2	64	2100	187000

If we have a sample of several independent observations (k -dimensional vectors) from our multivariate normal distribution, we can compute λ_0 for the whole sample by multiplying together the separate λ_0 's. This method may be regarded as an alternative to Hotelling's generalised "Student" test. (See, for example, Wilks [8], Section 11.4, where further references are given.)

REFERENCES

- [1] M. S. BARTLETT, "The statistical significance of odd bits of information," *Biometrika* Vol. 39 (1952), pp. 328-337.
- [2] I. J. GOOD, *Probability and the Weighing of Evidence*, Charles Griffin, London, 1950.
- [3] I. J. GOOD, "The appropriate mathematical tools for describing and measuring uncertainty," *Uncertainty and Business Decisions*, C. F. Carter, G. P. Meredith, and G. L. S. Shekelle (eds.), University Press, Liverpool, 1954.
- [4] I. J. GOOD, "A new finite series for Legendre polynomials," *Proc. Cam. Phil. Soc.*, Vol. 51 (1955), pp. 385-388.
- [5] H. JEFFREYS, *Theory of Probability*, Oxford University Press, 1939.
- [6] R. M. REDHEFFER, "A note on the surprise index," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 128-130.
- [7] W. WEAVER, "Probability, rarity, interest and surprise," *Scientific Monthly*, Vol. 67 (1948), pp. 390-392.
- [8] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1946.

DISTRIBUTIONS OF ROOTS OF QUADRATIC EQUATIONS WITH RANDOM COEFFICIENTS¹

BY JOHN W. HAMBLEN

Oklahoma Agricultural and Mechanical College

General Problem. The problem under consideration is, given the joint p.d.f. of the coefficients of an algebraic equation which can be expressed in polynomial form, to determine the joint p.d.f. of the roots and their marginal p.d.f.'s. Complete results are obtainable for the quadratic.

1. Introduction. Consider an algebraic equation which can be written in polynomial form as

$$(1.1) \quad \eta^n - \xi_1 \eta^{n-1} + \xi_2 \eta^{n-2} - \cdots + (-1)^n \xi_n = 0,$$

where the coefficients, $\xi_i (i = 1, \dots, n)$, are real or complex random variables with a given joint p.d.f. The roots of (1.1), $\eta_i (i = 1, \dots, n)$, are random variables which have a p.d.f. that depends upon the p.d.f. of the coefficients. To obtain the joint p.d.f. of the η_i it is apparent that we must consider the two cases, when the coefficients are real and complex, separately. Furthermore, when the coefficients are real the roots may be either real or complex and hence require separate treatment. The case where the ξ_i are complex random variables was considered in a note by M. A. Girshick [1]. When the ξ_i are real, the η_i may be real or complex. For real η_i the functional form of their p.d.f. is obtained by a change of variables in the p.d.f. of the ξ_i by the use of the relationships

$$(1.2) \quad \xi_1 = \sum_{i=1}^n \eta_i, \quad \xi_2 = \sum_{i < j} \eta_i \cdot \eta_j, \dots, \xi_n = \prod_{i=1}^n \eta_i,$$

with Jacobian, J , given by $\prod_{i < j} (\eta_i - \eta_j)$. For complex η_i the treatment is similar, but a new set of relationships must be found to replace (1.2). In this case, we must be able to express the ξ_i as functions of the real and imaginary parts of the η_i separately.

2. Limitations. We can now see that there are two major problems involved in determining the p.d.f. of the roots of (1.1) explicitly. The functional form of the p.d.f. can be obtained without difficulty. However, we must be able to determine what regions of the coefficient space will give rise to real roots and what regions will give complex roots. Secondly, after having identified these regions we must be able to define their transforms into the root space. At present, complete results are obtainable only for the quadratic.

3. Quadratic. For $n = 2$ we have

$$(3.1) \quad \eta^2 - \xi_1 \eta + \xi_2 = 0,$$

Received August 29, 1955; revised February 16, 1956.

¹ Part of a Ph.D. thesis presented to the Department of Mathematics, Purdue University, June, 1955. Work done under Purdue Research Foundation XR Fellowship Grant No. 1071.

where ξ_1 and ξ_2 are real random variables and hence may be any real-valued, Borel-measurable functions of real random variables.

The roots, η_1 and η_2 , of (3.1) are random variables associated with ξ_1 and ξ_2 by the relationships

$$(3.2) \quad \eta_1 = \frac{\xi_1}{2} + \sqrt{\frac{\xi_1^2}{4} - \xi_2}, \quad \eta_2 = \frac{\xi_1}{2} - \sqrt{\frac{\xi_1^2}{4} - \xi_2}$$

or

$$(3.3) \quad \xi_1 = \eta_1 + \eta_2, \quad \xi_2 = \eta_1 \cdot \eta_2$$

η_1 and η_2 are either both real or are complex conjugates. From (3.2) we see at once that all points belonging to the "interior" of the parabola $\xi_2 = \xi_1^2/4$ will give complex roots, while the remainder of the (ξ_1, ξ_2) -plane, which consists of points on and "outside" of the parabola, will give real roots.

We now consider the joint p.d.f., $f(x, y)$, of ξ_1 and ξ_2 , where $f(x, y)$ is of the continuous type. By truncating along the parabola $\xi_2 = \xi_1^2/4$, we obtain conditional p.d.f.'s relative to the hypotheses $\xi_2 > \xi_1^2/4$ and $\xi_2 \leq \xi_1^2/4$. If we let $P(R) = P(\xi_2 \leq \xi_1^2/4)$ and $P(C) = P(\xi_2 > \xi_1^2/4)$, then $P(R)$ and $P(C)$ are the probabilities of real and complex roots, respectively, and are given by

$$P(R) = \iint_{y \leq x^2/4} f(x, y) dy dx \quad \text{and} \quad P(C) = \iint_{y > x^2/4} f(x, y) dy dx.$$

The conditional or truncated p.d.f.'s [2] are

$$f(x, y | C) = f(x, y)/P(C), y > \frac{x^2}{4}; \quad f(x, y | R) = f(x, y)/P(R), y \leq \frac{x^2}{4}.$$

For $\xi_2 \leq \xi_1^2/4$, the roots of (3.1) are real and have a joint p.d.f. which is uniquely determined by the p.d.f. of the coefficients ξ_1 and ξ_2 . We will let $g(v_1, v_2 | R)$ denote the p.d.f. of the real roots. The functions (3.2) and (3.3) satisfy the sufficient conditions given by Cramér [2] for a change of variables in a continuous type density function. Therefore, we have

$$g(v_1, v_2 | R) = f(v_1 + v_2, v_1 v_2) |J| / P(R)$$

for all $v_1 \geq v_2$, where $|J| = (v_1 - v_2)$.

Let $g_1(v_1 | R)$ and $g_2(v_2 | R)$ be the marginal density functions of the real roots η_1 and η_2 , respectively. These are given by

$$g_1(v_1 | R) = \int_{-\infty}^{v_1} g(v_1, v_2 | R) dv_2, \quad \text{and} \quad g_2(v_2 | R) = \int_{v_2}^{\infty} g(v_1, v_2 | R) dv_1.$$

For $\xi_2 > \xi_1^2/4$, the roots of (3.1) are complex conjugates. Let $\eta_1 = \alpha + \beta i$, then $\eta_2 = \alpha - \beta i$. α and β are defined by the functions

$$(3.4) \quad \alpha = \xi_1/2, \quad \beta = \sqrt{\xi_2 - \frac{\xi_1^2}{4}},$$

or

$$\xi_1 = 2\alpha, \quad \xi_2 = \alpha^2 + \beta^2.$$

α and β have a joint p.d.f. which is uniquely determined by the p.d.f. of ξ_1 and ξ_2 , $f(x, y | C)$. Let $h_1(X, Z | C)$ denote the p.d.f. of α and β . The functions (3.4) satisfy the conditions stated by Cramér [2], so that we may find $h_1(X, Z | C)$ by a change of variables in $f(x, y | C)$. Therefore, we have

$$h_1(X, Z | C) = f(2X, X^2 + Z^2) |J| / P(C)$$

for all X and all $Z > 0$, where $|J| = 4Z$.

Similarly, if we let $h_2(X, Z | C)$ be the joint p.d.f. of α and $-\beta$, we will have

$$h_2(X, Z | C) = f(2X, X^2 + Z^2) |J'| / P(C)$$

for all X and all $Z < 0$, where $|J'| = -4Z$.

4. Examples. Numerous cases were considered [3] to the extent of expressing the marginal p.d.f.'s as integrals. For these cases ξ_1 and ξ_2 were categorized according to type of interval over which their p.d.f. was greater than zero. There are twelve different interval types as follows: $(-\infty, \infty)$, $(0, \infty)$, $(-\infty, 0)$, (A, ∞) , $(-\infty, -A)$, $(-A, \infty)$, $(-\infty, A)$, (A, B) , $(-A, -B)$, $(-A, B)$, $(0, A)$, $(-A, 0)$, where $A > 0$, $B > 0$. The various combinations were considered using the normal, gamma, and rectangular density functions, respectively, and assuming independence for ξ_1 and ξ_2 for convenience in obtaining their joint p.d.f.'s. Some dependent cases were also considered.

4.1. Example. Bivariate Normal. Let $f(x, y)$ be the general bivariate normal p.d.f., $n(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Then

$$(4.1.1) \quad g(v_1, v_2 | R) = \frac{(v_1 - v_2)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}P(R)} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{v_1 + v_2 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{v_1 + v_2 - \mu_1}{\sigma_1} \right) \left(\frac{v_1 v_2 - \mu_2}{\sigma_2} \right) + \left(\frac{v_1 v_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}$$

$$-\infty < v_2 \leq v_1, \quad -\infty < v_1 < \infty,$$

where

$$P(R) = \int_{-\infty}^{\infty} \int_{-\infty}^{x^2/4} n(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) dy dx.$$

If we let $u = (x - \mu_1) / \sigma_1$ and $w = (y - \mu_2) / \sigma_2$ we have

$$P(R) = \int_{-\infty}^{\infty} \int_{-\infty}^{\theta'(u)} n(u, w; \rho) dw du,$$

where

$$\theta'(u) = \frac{1}{4\sigma_2} [(\sigma_1 u + \mu_1)^2 - 4\mu_2].$$

On completing the square on w in the exponent, and substituting

$$t = \frac{w - u\rho}{\sqrt{1 - \rho^2}}, \quad dt = \frac{dw}{\sqrt{1 - \rho^2}},$$

we obtain

$$P(R) = \int_{-\infty}^{\infty} \int_{-\infty}^{\theta(u)} (2\pi)^{-1} \exp \left\{ -\frac{1}{2}(t^2 + u^2) \right\} dt du,$$

where $\theta(u) = [\theta'(u) - \rho u] / \sqrt{1 - \rho^2}$. Finally, we may write

$$(4.1.2) \quad P(R) = \int_{-\infty}^{\infty} \varphi(u) \Phi[\theta(u)] du,$$

where

$$\Phi(z) = \int_{-\infty}^z \varphi(t) dt$$

is the cumulative normal probability function and $\varphi(t)$ is the standardized normal density function. We must employ numerical methods of integration to find the value of $P(R)$ for a given $n(x, y)$, i.e., for a given set of values for $\mu_1, \mu_2, \sigma_1, \sigma_2$, and ρ .

For the complex roots we have

$$h_1(X, Z | C) = \frac{4Z}{P(C)} n(2X, X^2 + Z^2), \quad Z > 0,$$

and

$$h_2(X, Z | C) = h_1(X, -Z | C), \quad Z < 0,$$

where $P(C) = 1 - P(R)$.

The marginal distributions of the real roots have for density functions

$$(4.1.3) \quad g_1(v_1 | R) = \int_{-\infty}^{\infty} g(v_1, v_2 | R) dv_2, \quad -\infty < v_1 < \infty,$$

$$(4.1.4) \quad g_2(v_2 | R) = \int_{-\infty}^{\infty} g(v_1, v_2 | R) dv_1, \quad -\infty < v_2 < \infty.$$

On substituting (4.1.1) in (4.1.3), expanding the terms in the exponent, and collecting terms w.r.t. powers of v_2 , we obtain

$$(4.1.5) \quad g_1(v_1 | R) = \int_{-\infty}^{\infty} \frac{(v_1 - v_2)}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}P(R)} \cdot \exp \left\{ -\frac{1}{2(1 - \rho^2)} [m_1^2(v_1) \cdot v_2^2 - 2m_1(v_1) \cdot m_2(v_1) \cdot v_2 + m_2(v_1)] \right\} dv_2,$$

where

$$m_1^2(v_1) = (v_1 / \sigma_2 - \rho / \sigma_1)^2 + (1 - \rho^2) / \sigma_1^2,$$

TABLE 1

$\rho = 0, \pm.2, \pm.4, \pm.6, \pm.8, \pm.9$			
μ_1	μ_2	σ_1	σ_2
0	0	1	1
3	10	1	2
10	10	1	1
3	3	1	1
10	3	2	1
-10	3	2	1

TABLE 2

$\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$	
ρ	$P(R)$
.9	.5237 449
.8	.5453 219
.6	.5698 161
.4	.5872 947
.2	.5873 160
0	.5890 214
-.2	.5873 160
-.4	.5872 947
-.6	.5698 161
-.8	.5453 219
-.9	.5237 449

$$m_1(v_1) \cdot m_2(v_1) = \rho v_1^2 / \sigma_1 \sigma_2 - (1 / \sigma_1^2 + \rho \mu_1 / \sigma_1 \sigma_2 - \mu_2 / \sigma_2^2) v_1 + (\mu_1 / \sigma_1^2 - \rho \mu_2 / \sigma_1 \sigma_2),$$

and

$$m_3(v_1) = (v_1 / \sigma_1 - \mu_1 / \sigma_1 + \rho \mu_2 / \sigma_2)^2 + (1 - \rho^2) \mu_2^2 / \sigma_2^2.$$

If we carry out the same procedure on (4.1.4) w.r.t. v_1 , we arrive at

$$(4.1.6) \quad g_2(v_2 | R) = \int_{v_1}^{\infty} \frac{(v_1 - v_2)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}P(R)} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} [m_1^2(v_2) \cdot v_1^2 - 2m_1(v_2) \cdot m_2(v_2) \cdot v_1 + m_3(v_2)] \right\} dv_1.$$

Equations (4.1.2), (4.1.5), and (4.1.6) were evaluated for the various sets of the parameters shown in Table 1 using the ElectroData digital computer of the Statistical Laboratory, Purdue University [3]. Table 2 gives the values of $P(R)$ for the case $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$; a few representative graphs of the marginal p.d.f.'s, $g_1(v_1 | R)$ are shown in Fig. 1 for the same case. The curves for $g_2(v_2 | R)$ are mirror images of those for g_1 , the symmetry being due to the fact that $g_1(v_1 | R; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = g_2(-v_2 | R; -\mu_1, \mu_2, \sigma_1, \sigma_2, -\rho)$ and $v_1 = -v_2$, since $n(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = n(-x, y; -\mu_1, \mu_2, \sigma_1, \sigma_2, -\rho)$. The tails of the g_2 curves are shown as dashed lines in Fig. 1.

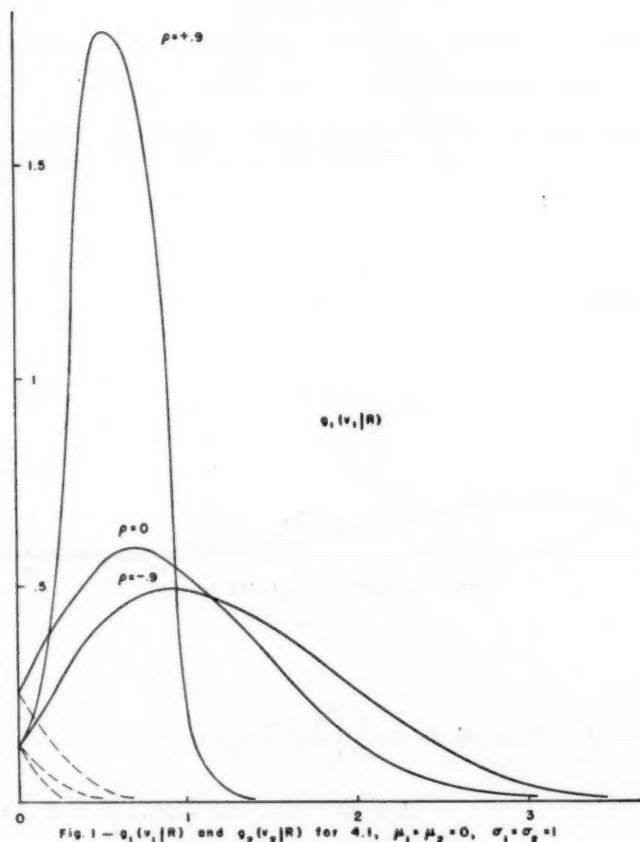


FIG. 1

4.2 Example. A Gamma Type. Let $f(x, y) = \exp \{-x - y\}$, $x \geq 0$, $y \geq 0$. Then

$$P(R) = \int_0^\infty \int_0^{x^{3/4}} \exp \{-x - y\} dy dx = 1 - 2e\sqrt{\pi}[1 - \Phi(\sqrt{2})] \doteq .24,$$

and

$$g(v_1, v_2 | R) = \frac{(v_1 - v_2)}{.24} \exp \{-(v_1 + v_2 + v_1 v_2)\}, \quad 0 \leq v_2 \leq v_1, 0 \leq v_1 \leq \infty;$$

$$h_1(X, Z | C) = \frac{4Z}{.76} \exp \{-(2X + X^2 + Z^2)\}, \quad X \geq 0, Z > 0;$$

$$h_2(X, Z | C) = \frac{-4Z}{.76} \exp \{-(2X + X^2 + Z^2)\}, \quad X \geq 0, Z < 0.$$

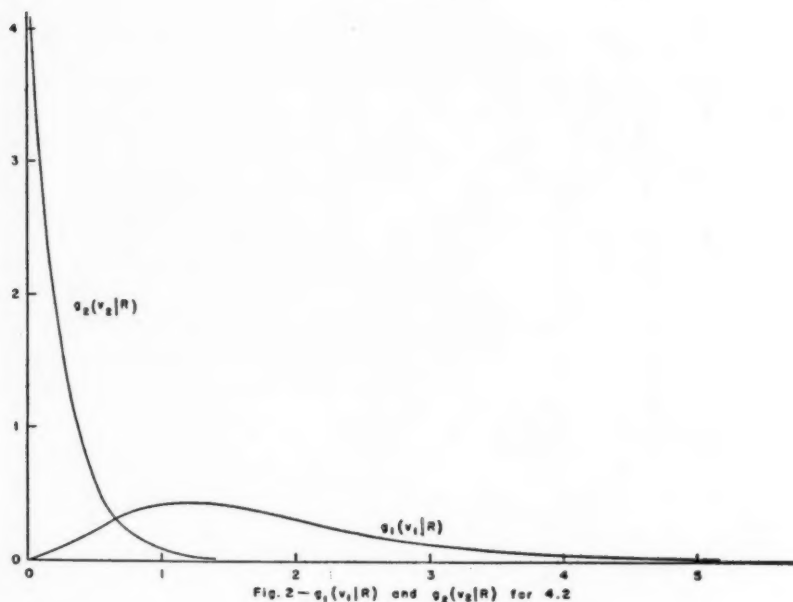
FIG. 2— $g_1(v_1|R)$ and $g_2(v_2|R)$ for 4.2

FIG. 2

Finally,

$$g_1(v_1 | R) = \int_0^{v_1} g(v_1, v_2 | R) dv_2 = \frac{1}{.24} \left[\frac{v_1^2 + v_1 - 1}{(1 + v_1)^2} \exp \{-v_1\} \right. \\ \left. + (1 + v_1)^{-2} \exp \{-(v_1^2 + 2v_1)\} \right], \quad 0 \leq v_1 < \infty,$$

and

$$g_2(v_2 | R) = \int_{v_2}^{\infty} g(v_1, v_2 | R) dv_1 = \\ \frac{1}{.24} (1 + v_2)^{-2} \exp \{-(v_2^2 + 2v_2)\}, \quad 0 \leq v_2 < \infty.$$

The frequency curves, $g_1(v_1 | R)$ and $g_2(v_2 | R)$, are plotted in Fig. 2.

5. Acknowledgments. The author is grateful to his major professor, Irving W. Burr, who suggested the problem, for his guidance and inspiration, and to the Statistical Laboratory of Purdue University for the use of their digital computer.

REFERENCES

- [1] M. A. GIRSHICK, "Note on the distribution of roots of a polynomial with random complex coefficients," *Ann. Math. Stat.*, Vol. 13 (1942), p. 235.
- [2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.
- [3] J. W. HAMBLIN, "Distributions of Roots of Algebraic Equations with Variable Coefficients," Ph.D. Thesis, Purdue University, 1955.

NOTES

A NOTE ON WEIGHTED RANDOMIZATION¹

By D. R. Cox²

Princeton University

Summary. It is shown that in simple statistical designs in which a covariance adjustment is made for concomitant variation, an unbiased between-treatment mean square can be produced by weighted randomization, i.e., by selecting an arrangement at random from a set of arrangements, giving different arrangements in the set unequal chances of selection.

1. Introduction. Randomization is one of the key elements in the statistical aspects of experimental design [2]. It has as its object the conversion, under rather weak assumptions, of uncontrolled variation of whatever form into effectively random variation. It thus makes the conclusions drawn from the experiment more objective and avoids the introduction of strong, and quite often unrealistic, assumptions about the uncontrolled variation. The methods of randomization in practical use depend on selecting one arrangement from a set giving each arrangement in the set equal chance of selection. For example in a randomized block design, the set would usually be that of all randomized block designs obtained by permuting the treatments within the chosen grouping of units into blocks. An arrangement for use would be one such design selected at random out of the set. The purpose of the present note is to point out the theoretical advantage in certain cases of choosing from the set with unequal probabilities. No recommendation is made about what should be done in practice in such situations.

The following assumption will be made throughout. We have N experimental units (plots, animals, etc.) and τ alternative treatments to be compared, one treatment being applied to each unit. Suppose that there is a quantity z_i associated with the i th unit and a constant a_μ associated with the μ th treatment, such that if the μ th treatment is applied to the i th unit, the resulting observation will be

$$(1) \quad z_i + a_\mu,$$

independently of the particular allocation of treatments to the other units. The z_i , a_μ are indeterminate to within a constant. The object of the experiment is considered to be the estimation, and possibly significance-testing, of linear contrasts among the a_μ . Assumption (1) can easily be generalized without

Received August 19, 1955.

¹ Research supported in part by the Office of Naval Research.

² Present address, Birkbeck College, University of London.

affecting the arguments that follow; for example a completely random term of constant mean and variance can be added to (1), but this will not be done here.

The essential points of (1) are that we are measuring on a scale on which treatment and unit terms are additive, that the treatment effects are constant, and that there is no competition or interference between different units.

The design is called unbiased [6] if it is possible to calculate from the observations the following:

- (i) unbiased estimates of the linear contrasts among the a_μ , for example of the differences, $a_\mu - a_\nu$;
- (ii) unbiased estimates of the variances of the estimates in (i);
- (iii) a mean square between treatments, s_b^2 , and the mean square for residual, s_r^2 , such that the expectation of s_b^2 is greater than or equal to that of s_r^2 , with equality if and only if $a_1 \cdots = a_r$.

It is well known that most of the standard designs are unbiased under Assumption (1); the quantities in (i)–(iii) are calculated by the usual analysis of variance methods, and expectations are taken over the set of arrangements from which the one actually used has been selected at random. We shall assume for the purpose of this paper that one requirement for a design to be satisfactory is that it should be unbiased.

Four examples of methods of design that are not unbiased in this sense under simple randomization are

- (a) a completely randomized, or other simple design, in which adjustments for a concomitant variable are made by analysis of covariance [5];
- (b) the so-called semi-Latin square [6];
- (c) a Latin square type cross-over design in which Assumption (1) is extended to allow for simple carry-over of treatment effects from one period to the next [5], [6];
- (d) certain designs in which for some practical reason there is a severe restriction on the treatment arrangements that are admissible. An example is [1].

This note is concerned with (a).

2. Adjustment for a concomitant variable. Suppose that on each experimental unit, a concomitant variable is measured, giving a value x_i for the i th unit, and that x_1, \dots, x_N are fixed and independent of the allocation of treatments to units. Nothing is assumed in the randomization analysis about the relation between the x 's and the z 's.

Consider to begin with a completely randomized experiment with n units for each treatment, $N = n\tau$. Denote the main observations, given by (1), by y_1, \dots, y_N . These are random variables depending on the particular arrangement of treatments selected. Let \sum_μ denote summation over those units receiving the μ th treatment. In the usual way define

$$(2) \quad \begin{aligned} \bar{y}_\mu &= \sum_\mu y_i / n, & \bar{x}_\mu &= \sum_\mu x_i / n, \\ \bar{y}_\cdot &= \sum y_i / N, & \bar{x}_\cdot &= \sum x_i / N, \end{aligned}$$

and set up the analysis of covariance table

	x^2	xy	y^2	d.f.
Between treatments	B_{xx}	B_{xy}	B_{yy}	$\tau - 1$
(3) Residual	R_{xx}	R_{xy}	R_{yy}	$\tau(n - 1)$
Total	T_{xx}	T_{xy}	T_{yy}	$\tau n - 1$

where, for example,

$$(4) \quad R_{xy} = \sum_{j=1}^{\tau} \sum_{i=1}^n (x_i - \bar{x}_j)(y_i - \bar{y}_j).$$

Let $b_r = R_{xy} / R_{xx}$ and define the adjusted treatment means

$$(5) \quad \hat{y}_\mu = \bar{y}_\mu - b_r(\bar{x}_\mu - \bar{x}).$$

Also define an estimate of the variance of $\hat{y}_\mu - \hat{y}_\nu$ by

$$(6) \quad V_s(\hat{y}_\mu - \hat{y}_\nu) = s_r^2 \left\{ \frac{2}{n} + \frac{(\bar{x}_\mu - \bar{x}_\nu)^2}{R_{xx}} \right\},$$

with

$$(7) \quad \text{Av}_{\mu, \nu} V_s(\hat{y}_\mu - \hat{y}_\nu) = s_r^2 \left\{ \frac{2}{n} + \frac{2}{n(\tau - 1)} \frac{B_{xx}}{R_{xx}} \right\},$$

where

$$(8) \quad s_r^2 = \frac{1}{(n\tau - \tau - 1)} (R_{yy} - R_{xy}^2/R_{xx})$$

is the residual mean square of y adjusting for regression on x . Finally the mean square for treatments adjusting for regression on x is

$$(9) \quad s_b^2 = \frac{1}{(\tau - 1)} \left(T_{yy} - \frac{T_{xy}^2}{T_{xx}} - R_{yy} + \frac{R_{xy}^2}{R_{xx}} \right).$$

The definition of these quantities is based on the least-squares theory of analysis of covariance; we now consider the randomization theory.

3. A method for calculating expectations under randomization. To investigate randomization expectations an elegant and powerful method due to Grundy and Healy [3] will be used. Denote the expectation under simple (unweighted) randomization of any function, f , of the observations by $E_r(f)$,

$$(10) \quad E_r(f) = \frac{1}{(\text{no. of possible arrangements})} \times \sum_{\text{all arrangements}} f.$$

Consider, as an example of the method, the calculation of $E_r(R_{yy})$, when the design is completely randomized. It is easily seen from (1) that this expectation is independent of a_1, \dots, a_τ and is a homogeneous completely symmetric function of z_1, \dots, z_N of degree two, invariant under translation of the z 's.

Therefore

$$(11) \quad E_F(R_{yy}) = \alpha \sum_{i=1}^N (z_i - \bar{z})^2,$$

identically in z_1, \dots, z_N , where α is a constant. If z_1, \dots, z_N is a random sample from a population of variance σ^2 , the expectations of the left- and right-hand sides of (11) are respectively $\tau(n-1)$ and $\alpha(\tau n-1)$, whence $\alpha = \tau(n-1)/(\tau n-1)$. Thus

$$(12) \quad E_F(R_{yy}) = \frac{\tau(n-1)}{\tau n-1} \sum_{i=1}^N (z_i - \bar{z})^2.$$

The choice of z_1, \dots, z_N as a random sample in the last step of the argument has no physical significance and is purely a mathematical device to exploit knowledge of the behavior of R_{yy} under the usual hypotheses of least-squares theory; see also [4].

In general the method is to establish the general form of the expectation by considerations of symmetry and invariance and then to find the precise expression by special choice of the z 's, exploiting our knowledge of what happens under the conditions of least-squares theory. Thus suppose that we require to show that for a Latin square the expectations of the mean squares for treatments and residual are equal, when there are no treatment effects. Consideration of symmetry and invariance show that both expectations are multiples of the residual sum of squares of the z 's, considered as a row \times column arrangement. Equality of the two expectations under least-squares theory then proves this equality under general randomization theory.

The result (12) is well known and can be obtained directly without difficulty. The point of Grundy and Healy's method is that it avoids enumerative calculations, and its advantage is consequently greater in the more complicated situations, such as, for example, in the proofs that Latin squares, balanced incomplete blocks, and so on are unbiased under (1).

4. The application to covariance adjustments. If we try to calculate the randomization expectations of s_r^2, s_s^2 , defined in (8) and (9), there is the difficulty that R_{xy}^2/R_{xx} is a ratio of random variables so that no simple exact expression for the form of its expectation can be written down. When $\alpha_1 = \dots = \alpha_r$, $T_{xx}, T_{xy} = T_{xx}, T_{yy} = T_{xx}$ are constant and it follows from (8), (9), and (12) that $E_F(s_r^2) = E_F(s_s^2)$ if and only if $E_F(R_{xy}^2/R_{xx})$ is linearly related in a particular way to T_{xx} and T_{xx}^2/T_{xx} . Consideration of the form of $E_F(R_{xy}^2/R_{xx})$ shows that no such relation can hold identically in the x 's and the z 's. Hence there is, in general, bias, although we always have

$$(13) \quad E_F(\hat{y}_\mu - \hat{y}_r) = a_\mu - a_r.$$

The bias arises from the factor $1/R_{xx}$ in (8) and so it is natural to try to remove the bias by weighting each arrangement of treatments proportionally

to R_{xx} . This is the general idea behind the following considerations. Suppose that the values x_1, \dots, x_N are available to the experimenter prior to the allocation of treatments to units. Let w be any non-negative function of x_1, \dots, x_N , defined for each arrangement of treatments within the set in which each treatment occurs n times. Let an arrangement be selected for use giving each design in the set a probability of selection proportional to w ; we shall call this a process of weighted randomization using w as weight function. If f is any function of the observations y and x , its expectation under weighted randomization is $E_w(f)$, where

$$(14) \quad E_w(f) = E_P(wf)/E_P(w).$$

Let $w = R_{xx}$ and consider $E_w(s_r^2)$. By (15) we need to know $E_P(R_{xx}R_{yy} - R_{xy}^2)$. This is independent of a_1, \dots, a_r and is homogeneous and of degree two in x_1, \dots, x_N and in z_1, \dots, z_N separately, and is unaffected by interchanging the x 's with the z 's. Hence

$$(15) \quad \begin{aligned} E_P(R_{xx}R_{yy} - R_{xy}^2) &= A\bar{x}^2\bar{z}^2 \\ &\quad + B(\bar{x}^2T_{xx} + \bar{z}^2T_{zz}) \\ &\quad + CT_{xx}T_{zz} + DT_{xx}^2 \\ &\quad + E \sum_{i=1}^N (x_i - \bar{x})(z_i - \bar{z})^2 \\ &\quad + F\bar{x}\bar{z}T_{xx} \\ &\quad + G \left\{ \bar{x} \sum_{i=1}^N (x_i - \bar{x})(z_i - \bar{z})^2 + \bar{z} \sum_{i=1}^N (x_i - \bar{x})^2(z_i - \bar{z}) \right\}, \end{aligned}$$

where A, \dots, G are constants. The simplest way of verifying (15) from first principles is to note that

$$\begin{aligned} \sum x_i^2 z_i^2, \quad \sum (x_i^2 z_i z_j + x_i x_j z_i^2), \quad \sum (x_i^2 z_k z_l + x_k x_l z_i^2), \\ \sum x_i x_j z_i z_j, \quad \sum x_i x_j z_i z_k, \quad \sum x_i x_j z_k z_l \end{aligned}$$

are the seven types of sum with the requisite degree of symmetry and that the right-hand side of (15) has seven arbitrary constants. $R_{xx}R_{yy} - R_{xy}^2$ is unaffected by changing x_i to $x_i + a$ and z_i to $z_i + b$, $i = 1, \dots, N$. Since this is true identically in the x 's and z 's, $A = B = F = G = 0$. Next, if $z_i = \lambda x_i$, $i = 1, \dots, N$, $R_{xx}R_{yy} - R_{xy}^2$ is identically zero, so that

$$(16) \quad 0 = \lambda^2 \left\{ CT_{xx}^2 + DT_{xx}^2 + E \sum_{i=1}^N (x_i - \bar{x})^4 \right\},$$

whence $E = 0$. $C = -D$. If we combine this result with the expression for $E_P(R_{xx})$ corresponding to (12), we have

$$(17) \quad E_w(s_r^2) = H(T_{xx} - T_{xx}^2 / T_{xx}),$$

where H is a constant. Finally let x_1, \dots, x_N be arbitrary but fixed and let the z_i be uncorrelated random variables with means βx_i and constant variance σ^2 . If E denotes expectation over their distribution,

$$E(T_{zz} - T_{zz}^2 / T_{zz}) = (n\tau - 2)\sigma^2, \\ E(s_r^2) = \sigma^2,$$

by the ordinary theory of regression. Finally, since $EE_w(s_r^2) = E_w E(s_r^2)$, (17) leads to $H = 1/(n\tau - 2)$, so that

$$(18) \quad E_N(s_r^2) = \frac{1}{(n\tau - 2)} \left(T_{zz} - \frac{T_{zz}^2}{T_{zz}} \right) = \sigma_r^2,$$

say.

Similarly if $a_1 = \dots = a_r$,

$$(19) \quad E_w(s_b^2) = \sigma_r^2,$$

the unbiased property. When the a 's are not all equal, a multiple of their corrected sum of squares is added to (19); details will not be given.

We now investigate the corresponding theory for the variance and estimated variance of the difference between two adjusted treatment means, $\hat{y}_\mu - \hat{y}_\nu$, say. It does not seem possible to obtain exact results corresponding to those for s_b^2 and s_r^2 and we shall need to use the following asymptotic results. If, as N tends to infinity, f and g are random variables, functions of the y 's and the x 's with fixed means and with variance of order $1/N$, then

$$(20) \quad E(fg) = E(f)E(g) + O(1/N),$$

and, under weak conditions on g ,

$$(21) \quad E(f/g) = E(f) / E(g) + O(1/N).$$

These will be used with E standing for E_F or E_w as convenient. The expectation on the left of (21) is to be taken as referring to the asymptotic distribution of f/g .

Now we have from (7), (18), and (20) that

$$(22) \quad E_w[\text{Av}_{\mu, \nu} V_s(\hat{y}_\mu - \hat{y}_\nu)] = \frac{2\sigma_r^2}{n} + \frac{2\sigma_r^2}{n(\tau - 1)} E_w \left(\frac{B_{zz}}{R_{zz}} \right) \left(1 + O\left(\frac{1}{N}\right) \right)$$

$$(23) \quad = \frac{2\sigma_r^2}{n} \left\{ 1 + \frac{1}{\tau(n - 1)} \left(1 + O\left(\frac{1}{N}\right) \right) \right\}.$$

If τ is fixed and n tends to infinity, the relative error in (20) is of order $1/N^2$. In obtaining (22) we have assumed that the x 's and the z 's are such as to make the variances of s_r^2 and B_{zz} / R_{zz} of order $1/N$.

Similarly to find the actual variance of $\hat{y}_\mu - \hat{y}_\nu$, we have that

$$(24) \quad \text{Av}_{\mu, \nu} \{ (\hat{y}_\mu - \hat{y}_\nu) - (a_\mu - a_\nu) \}^2 = \text{Av}_{\mu, \nu} \left\{ \frac{1}{n} (\Sigma_\mu z - \Sigma_\nu z) - \frac{b_r}{n} (\Sigma_\mu x - \Sigma_\nu x) \right\}^2 \\ = \frac{2}{n(\tau - 1)} \left(B_{zz} - \frac{2R_{zz}B_{zz}}{R_{zz}} + \frac{R_{zz}^2 B_{zz}}{R_{zz}^2} \right).$$

If we apply the operator E_w to (24), we get the required variance. The expectation can be evaluated in a way similar to (23), dealing with the last term by (21). The final answer is the right-hand side of (23). That is, to the order indicated above, (7) is an unbiased estimate of the average variance of $\hat{y}_\mu - \hat{y}_\nu$.

5. Extensions. The calculations in Section 4 have, for simplicity, been made for the completely randomized design. However the results can be extended to designs such as randomized blocks and Latin squares; weighting proportional to the residual sum of squares of x again gives an unbiased treatment mean square. Another generalization is to multiple analysis of covariance, in which the treatment means are adjusted for k concomitant variables x_1, \dots, x_k . The appropriate weighting function is then the residual generalized variance, i.e., the determinant $|R_{ij}|$, where R_{ij} is the residual sum of products of x_i and x_j .

6. Discussion. The idea of weighted randomization discussed above is probably solely of theoretical interest, at any rate in the context considered here. A full discussion of possible practical applications would require further work, but the following points are worth making.

(i) The bias in unweighted randomization is probably small, except possibly when N is very small and the correlation between the z 's and the x 's very non-linear. Further work is needed, however, to find the likely magnitude of the bias in typical cases.

(ii) Weighted randomization is perhaps most likely to be of practical value when a series of similar experiments are planned, each with a small value of N . Another possible application is to Latin square designs in which it is desired to control variations diagonally across the square, in addition to row and column variation. This can be done by inserting a suitable concomitant variable, for example the product of row number and column number suitably coded. Weighted randomization would justify such a method in the same way that ordinary randomization justifies the conventional use of the Latin square.

(iii) Arrangements with a large value of R_{xx} will have a small value for B_{xx} and conversely. Hence the weighting proportional to R_{xx} attaches greater chance of selection to those arrangements in which the treatment groups are balanced with respect to the mean value of x .

(iv) If weighted randomization is to be done in practice with N not very small, some short-cut method is needed for selecting an arrangement, since the enumeration of all arrangements and the calculation of R_{xx} for each would usually be too tedious. Professor J. W. Tukey has pointed out that weighted randomization can be done reasonably simply as follows. Let M be the maximum over-all arrangement of R_{xx} . Select an arrangement by unweighted randomization and calculate R_{xx} for it. Reject the arrangement with probability $1 - R_{xx} / M$. Continue until an arrangement is accepted.

(v) Weighted randomization is, of course, restricted to cases in which the concomitant variable is available prior to the allocation of treatments to units.

REFERENCES

- [1] D. R. COX, "The design of an experiment in which certain treatment arrangements are inadmissible," *Biometrika*, Vol. 41 (1954), pp. 287-295.
- [2] R. A. FISHER, *The Design of Experiments*, 5th ed., Oliver and Boyd, Ltd., 1949.
- [3] P. M. GRUNDY AND M. J. R. HEALY, "Restricted randomization and quasi-Latin squares," *J. Roy. Stat. Soc. (B)*, Vol. 12 (1950), pp. 286-291.
- [4] J. O. IRWIN AND M. G. KENDALL, "Sampling moments of moments for a finite population," *Ann. Eugenics* Vol. 12 (1943-45), pp. 138-142.
- [5] O. KEMPTHORNE, *The Design and Analysis of Experiments*, J. Wiley and Sons, 1952.
- [6] F. YATES, "Bases logiques de la planification des expériences," *Ann. Inst. H. Poincaré* Vol. 12 (1951), pp. 97-112.

ON STOCHASTIC APPROXIMATION METHODS¹

BY J. WOLFOWITZ

Cornell University

In [1] A. Dvoretzky proved the theorem quoted below, which implies all previous results on the convergence to a limit of stochastic approximation methods. (For a description of these results see [1].) In the present note we give a simple and, we think, perspicuous proof of this theorem which may be of help in further work. The present note is entirely self-contained and may be read without reference to [1].

THEOREM. (Dvoretzky) Let α_n , β_n and γ_n ($n = 1, 2, \dots$) be non-negative real numbers satisfying

$$(1) \quad \lim_{n \rightarrow \infty} \alpha_n = 0,$$

$$(2) \quad \sum_{n=1}^{\infty} \beta_n < \infty,$$

and

$$(3) \quad \sum_{n=1}^{\infty} \gamma_n = \infty.$$

Let θ be a real number and T_n ($n = 1, 2, \dots$) be measurable transformations satisfying

$$(4) \quad |T_n(r_1, \dots, r_n) - \theta| \leq \max[\alpha_n, (1 + \beta_n)|r_n - \theta| - \gamma_n]$$

for all real r_1, \dots, r_n . Let X_1 and Y_n ($n = 1, 2, \dots$) be random variables and define²

Received December 16, 1955.

¹ This research was supported by the United States Air Force under Contract No. AF18-(600)-685 monitored by the Office of Scientific Research.

² In the proof of the theorem we will, for the sake of brevity, write $T_n(X_n)$ for

$$T_n(X_1, \dots, X_n),$$

just as is done in [1]. No ambiguity will be caused by this.

$$(5) \quad X_{n+1}(\omega) = T_n(X_1(\omega), \dots, X_n(\omega)) + Y_n(\omega)$$

for $n \geq 1$.

Then the conditions $E\{X_1^2\} < \infty$,

$$(6) \quad \sum_{n=1}^{\infty} E\{Y_n^2\} < \infty$$

and

$$(7) \quad E\{Y_n | X_1, \dots, X_n\} = 0$$

with probability 1 for all n , imply

$$(8) \quad \lim_{n \rightarrow \infty} E\{(X_n - \theta)^2\} = 0$$

and

$$(9) \quad P\{\lim_{n \rightarrow \infty} X_n = \theta\} = 1.$$

EXTENSION. The theorem remains valid if α_n and β_n in (4) are replaced by non-negative functions $\alpha_n(r_1, \dots, r_n)$ and $\beta_n(r_1, \dots, r_n)$ respectively, provided: The functions $\alpha_n(r_1, \dots, r_n)$ are uniformly bounded and

$$(10) \quad \lim_{n \rightarrow \infty} \alpha_n(r_1, \dots, r_n) = 0$$

uniformly for all sequences r_1, \dots, r_n, \dots ; the functions $\beta_n(r_1, \dots, r_n)$ are measurable and

$$(11) \quad \sum_{n=1}^{\infty} \beta_n(r_1, \dots, r_n)$$

is uniformly bounded and uniformly convergent for all sequences r_1, \dots, r_n, \dots ; and for any $L > 0$ there exist non-negative functions $\gamma_n(r_1, \dots, r_n)$ satisfying (4), and

$$(12) \quad \sum_{n=1}^{\infty} \gamma_n(r_1, \dots, r_n) = \infty$$

holds uniformly for all sequences r_1, \dots, r_n, \dots for which

$$(13) \quad \sup_{n=1,2,\dots} |r_n| < L.$$

PROOF: Without loss of generality we may take $\theta = 0$.

I. From (4) and (6) it follows readily that $EX_n^2 < \infty$ for any n .

II. Define $s(n)$ to be the sign of $[T_n(X_n)][X_n]$ if neither factor is zero, and $s(n) = 1$ if either factor is zero. Define $\pi(m, n) = \prod_{j=m}^n s(j)$, $Y'_n = \pi(1, n)Y_n$. The series $\sum_1^{\infty} Y'_n$ converges w.p.1, by Loève ([2], p. 387, D) and (6) and (7). Let

$$Z(m, n) = \sum_{j=m}^n Y'_j$$

For any δ and ϵ both >0 , there exists $M'(\delta, \epsilon)$ such that

$$(14) \quad P \left\{ \sup_{\substack{m, n \\ M' \leq m \leq n}} |Z(m, n)| > \frac{\delta}{48} \right\} < \frac{\epsilon}{2}.$$

III. Let $d(m, m-1) = 1$ and, for $n \geq m$,

$$d(m, n) = \prod_{j=m}^n (1 + \beta_j).$$

Consider the sum

$$S(m, n) = \sum_{j=m}^{n+1} d(j, n) Y'_{j-1},$$

which is equal to

$$(15) \quad \sum_{j=m}^{n-1} Z((m-2), (j-1)) [d(j, n) - d(j+1, n)] \\ - Y'_{m-2} d(m, n) + Z((m-2), (n-1)) d(n, n) + Y'_n.$$

Since $d(j, n) \geq d(j+1, n)$ we have that the absolute value of (15) is not greater than

$$2 \left[\sup_{m-1 \leq j \leq n} |Z((m-2), (j-1))| (d(m, n)) + |Y_n| \right].$$

Hence, from (11) and (14) it follows that, for δ and ϵ both >0 , there exists an $M''(\delta, \epsilon) \geq M'(\delta, \epsilon)$ such that $d(m, \infty) < \frac{3}{2}$ for $m \geq M''$ and

$$(16) \quad P \left\{ \sup_{\substack{m, n \\ M'' \leq m \leq n}} |Z(m, n)| < \frac{\delta}{48}, \sup_{\substack{m, n \\ M'' \leq m \leq n}} |S(m, n)| < \frac{\delta}{8} \right\} > 1 - \frac{\epsilon}{2}.$$

Proof of (9) under the conditions of the extension. Let ϵ and δ be positive and arbitrary. It is sufficient to prove that

$$(17) \quad P\{|X_n| < \delta \text{ for all } n \text{ sufficiently large}\} > 1 - \epsilon.$$

Let $M \geq M''(\delta, \epsilon)$ be so large that, for $n \geq M$, $\alpha_n < \delta/8$. Let L be so large that $L > \delta$ and

$$(18) \quad \max_{1 \leq j \leq M} EX_j^2 < \frac{\epsilon L^2}{32M}.$$

We take this to be the L for which (12) holds. It also follows that

$$(19) \quad P \left\{ \max_{1 \leq j \leq M} |X_j| \leq \frac{L}{4} \right\} > 1 - \frac{\epsilon}{2}.$$

Suppose that the following four conditions are fulfilled:

$$(20) \quad \text{The relations in curly brackets in (16);}$$

$$(21) \quad |X_m| \leq \frac{\delta}{4} \text{ for some } m \geq M;$$

$$(22) \quad |X_{m+j}| > \frac{\delta}{4}, \quad 1 \leq j \leq k;$$

$$(23) \quad |X_{m+k+1}| \leq \frac{\delta}{4}.$$

Here $1 \leq k \leq \infty$. In case $k = \infty$, (22) is to hold for all $j \geq 1$ and (23) is vacuous. (It will be clear by the time the proof is finished that k cannot be ∞ .) Because $\alpha_n < \delta/8$ for $n \geq M$ and because of (20), (21), and (22) it follows that

$$(24) \quad |T_{m+j}(X_{m+j})| > \alpha_{m+j}, \quad 0 \leq j \leq k-1,$$

$$(25) \quad \text{sign } X_{m+j+1} = \text{sign } T_{m+j}(X_{m+j}), \quad 0 \leq j \leq k-1.$$

Applying (4) (with the γ 's zero) we obtain that X_{m+1} lies between zero and

$$(26) \quad s(m)(1 + \beta_m)X_m + Y_m.$$

Repeating this argument, we obtain that, for $1 \leq j \leq k$, X_{m+j} lies between 0 and

$$(27) \quad \begin{aligned} & s(m+j-1)s(m+j-2) \cdots s(m)d(m, m+j-1)X_m \\ & + s(m+j-1) \cdots s(m+1)d(m+1, m+j-1)Y_m + \cdots \\ & + s(m+j-1)d(m+j-1, m+j-1)Y_{m+j-2} + Y_{m+j-1}. \end{aligned}$$

The absolute value of (27) is not greater than

$$(28) \quad |X_m|d(m, m+j-1) + |S(m+1, m+j-1)|.$$

Hence

$$(29) \quad |X_{m+j}| < \delta, \quad 1 \leq j \leq k.$$

To prove (17) it remains only to show that the following conditions cannot both hold:

$$(30) \quad \text{the relations in curly brackets in (16) and (19);}$$

$$(31) \quad |X_n| > \frac{\delta}{4} \text{ for all } n \geq M.$$

Applying the argument of the previous paragraph with δ replaced by L we obtain that

$$(32) \quad |X_n| < L \text{ for all } n \geq 1.$$

Hence (12) holds. In view of (30) and (31) it follows that

$$(33) \quad |T(X_n)| > \alpha_n \text{ for all } n \geq M-1,$$

$$(34) \quad \text{sign } T_n(X_n) = \text{sign } X_{n+1} \text{ for all } n \geq M-1.$$

We may now, and do, apply the argument which led to (28), but with the γ 's which satisfy (12). We conclude that, for all $n > M$, the absolute value of $|X_n|$ is not greater than

$$(35) \quad |X_M| d(M, n-1) + |S(M+1, n-1)| - \sum_{j=M}^{n-1} \gamma_j$$

For n sufficiently large this becomes negative, contradicting (33) and hence (31). This completes the proof of (9).

The fact that $EX_1^2 < \infty$ is used in the above proof only in order that $EX_n^2 < \infty$ for all n , and this latter fact is needed only for (8), and not for (9). For in the proof above we used the fact that $EX_n^2 < \infty$ only to obtain explicitly an L for which (19) holds. Such an L obviously exists whether or not $EX_n^2 < \infty$.

Proof of (8) under the conditions of the extension. Let $K = \max_{1 \leq j < \infty} \alpha_j$. Let N be an integer to be chosen later. In view of (9) we have only to prove that $\lim_{n \rightarrow \infty} E\{(|X_n| - K)^+\}^2 = 0$. Let P denote probability measure and A be any set in the sample space which can be defined in terms of X_1, \dots, X_m . We use the inequality

$$(36) \quad \begin{aligned} H_{m+1}(A) &= \int_A ((|X_{m+1}| - K)^+)^2 dP = \int_A ((|T_m(X_m) + Y_m| - K)^+)^2 dP \\ &\leq \int_A [Y_m^2 + ((|T_m(X_m)| - K)^+)^2] dP \\ &\leq \int_A [Y_m^2 + K\beta_m(1 + K\beta_m) + (1 + \beta_m)^2(1 + K\beta_m)((|X_m| - K)^+)^2] dP \end{aligned}$$

which is in [1] and can be deduced from (4) and (7). Let $B(j)$ be the set $\{|X_{N+j}| \leq K, |X_{N+i}| > K \text{ for } 0 \leq i < j\}$, $D(j)$ the complement of

$$B(0) + B(1) + \dots + B(j).$$

Iterate the inequality (36) to obtain an upper bound on $H_n(A)$, $n > N$, beginning the iteration at $m = N, N+1, \dots, n-1$, respectively, and using as A the sets $B(0), B(1), \dots, B(n-N-1)$, respectively. In each case the last term of the integrand of the right member of (36) vanishes. Adding, we obtain that $H_n(B(0) + \dots + B(n-N))$ can be made arbitrarily small by making N sufficiently large.

It remains only to consider $H_n(D(n-N))$. For any point in $D(n-N)$ we have, as in (27), that

$$(37) \quad |X_n| \leq |\pi(1, N-1) d(N, n-1) X_N + S(N+1, n-1)|$$

Hence, by Minkowski's inequality

$$(38) \quad \begin{aligned} &\left(\int_{D(n-N)} (X_n)^2 dP \right)^{\frac{1}{2}} \\ &\leq [d(1, \infty)] \left(\int_{D(n-N)} (X_N)^2 dP \right)^{\frac{1}{2}} + [d(1, \infty)] \left(\sum_{j=N}^{\infty} EY_j^2 \right)^{\frac{1}{2}} \end{aligned}$$

The second term on the right of (38) can be made arbitrarily small by making N sufficiently large. The first term can be made arbitrarily small by making n sufficiently large, since $P\{D(n - N)\} \rightarrow 0$ as $n \rightarrow \infty$. This completes the proof of (8).

REFERENCES

- [1] A. DVORETZKY, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1956.
 [2] M. LOÈVE, *Probability Theory*, D. Van Nostrand Company, 1955.

ON THE DERIVATIVES OF A CHARACTERISTIC FUNCTION AT THE ORIGIN

BY E. J. G. PITMAN

University of Tasmania

1. Introduction. Let $F(x)$, $-\infty < x < \infty$, be a distribution function, and

$$\phi(t) = \int_{-\infty}^{\infty} e^{itz} dF(x)$$

its characteristic function, defined and continuous for all real t . Let k be a positive integer. If the k th moment of $F(x)$,

$$\mu_k = \int_{-\infty}^{\infty} x^k dF(x),$$

exists and is finite (integral absolutely convergent), $\phi(t)$ has a finite k th derivative for all real t given by

$$\phi^{(k)}(t) = i^k \int_{-\infty}^{\infty} x^k e^{itz} dF(x).$$

In particular,

$$\phi^{(k)}(0) = i^k \mu_k.$$

The existence and finiteness of μ_k is a sufficient condition for the existence and finiteness of $\phi^{(k)}(0)$. It can be shown (see [1]) that when k is even, this condition is also necessary; but when k is odd this is not so. Zygmund [2] has given a necessary and sufficient condition for the existence of $\phi'(0)$ and also one for the existence of a symmetric derivative of higher odd order at $t = 0$; but he imposes a certain condition (smoothness) on the characteristic function. In the following theorem the conditions are on the distribution function only.

Received October 5, 1955

2. Statement of Results.

THEOREM. Let k be an odd positive integer. Necessary and sufficient conditions for the existence of $\phi^{(k)}(0)$ are:

$$(i) \quad \lim_{x \rightarrow \infty} x^k \{F(-x) + 1 - F(x)\} = 0,$$

$$(ii) \quad \lim_{T \rightarrow \infty} \int_{-T}^T x^k dF(x) \text{ exists.}$$

When these two conditions are satisfied,

$$\phi^{(k)}(0) = i^k \lim_{T \rightarrow \infty} \int_{-T}^T x^k dF(x).$$

If X is a random variable with distribution function $F(x)$, so that

$$F(x) = P\{X \leq x\},$$

condition (i) may be stated in the form

$$\lim_{x \rightarrow \infty} x^k [P\{X \leq -x\} + P\{X > x\}] = 0.$$

A condition which is easily proved equivalent is

$$\lim_{x \rightarrow \infty} x^k \{P\{|X| \geq x\}\} = 0.$$

3. Two lemmas.

LEMMA 1. If $G(x)$ is defined and non-decreasing for $x \geq 0$, and if $k > 0$, the four statements below are equivalent, i.e., any one implies the other three.

$$(1) \quad \lim_{T \rightarrow \infty} T^k \int_T^\infty dG(x) = 0;$$

$$(2) \quad \lim_{T \rightarrow \infty} \frac{\int_0^T x^{k+1} dG(x)}{T} = 0;$$

$$(3) \quad \lim_{T \rightarrow \infty} T \int_T^\infty x^{k-1} dG(x) = 0;$$

$$(4) \quad \lim_{T \rightarrow \infty} T \int_0^\infty x^{k-1} \sin^2(x/T) dG(x) = 0.$$

Suppose (1) is true. Put

$$H(x) = \int_x^\infty dG(x) = G(\infty) - G(x).$$

Then $T^k H(T) \rightarrow 0$ when $T \rightarrow \infty$, and

$$\frac{\int_0^T x^{k+1} dG(x)}{T} = \frac{-\int_0^T x^{k+1} dH(x)}{T} = -T^k H(T) + \frac{(k+1) \int_0^T x^k H(x) dx}{T},$$

both terms of which $\rightarrow 0$ as $T \rightarrow \infty$ if $T^k H(T) \rightarrow 0$, and so (2) is true. Now

$$2(2T)^{-1} \int_0^{2T} x^{k+1} dG(x) \geq T^{-1} \int_T^{2T} x^{k+1} dG(x) \geq T^k \int_T^{2T} dG(x).$$

When (2) is true, the first term in the inequality $\rightarrow 0$ as $T \rightarrow \infty$, and therefore so does the last, i.e.,

$$(5) \quad W(T) = T^k \{G(2T) - G(T)\} \rightarrow 0 \text{ as } T \rightarrow \infty.$$

$$T^k \int_T^\infty dG(x) = T^k \sum_{n=1}^\infty \{G(2^n T) - G(2^{n-1} T)\} = \sum_{n=1}^\infty 2^{-(n-1)k} W(2^{n-1} T).$$

Because of (5), $W(T)$ is bounded for $T \geq 0$, and therefore this series is uniformly convergent with respect to $T \geq 0$. When $T \rightarrow \infty$, each term $\rightarrow 0$, and therefore (1) is true. Thus (2) implies (1).

Suppose again that (1) is true. Put

$$A(T) = \sup [x^k H(x); x \geq T].$$

Then $A(T) \rightarrow 0$ as $T \rightarrow \infty$, and

$$\begin{aligned} T \int_T^\infty x^{k-1} dG(x) &= -T \int_T^\infty x^{k-1} dH(x) \\ &= T^k H(T) + (k-1)T \int_T^\infty x^{k-2} H(x) dx \\ &\leq T^k H(T) + |k-1| TA(T) \int_T^\infty x^{-2} dx \\ &= T^k H(T) + |k-1| A(T), \end{aligned}$$

which $\rightarrow 0$ as $T \rightarrow \infty$. Thus (1) implies (3).

The converse of this is not actually used in this paper; but there is some interest in stating and proving it so as to round out the lemma. If $k \geq 1$,

$$T \int_T^\infty x^{k-1} dG(x) \geq T^k \int_T^\infty dG(x),$$

and so (3) implies (1) in this case. We now suppose $0 < k < 1$. Now

$$T \int_T^{2T} x^{k-1} dG(x) \geq 2^{k-1} T^k \int_T^{2T} dG(x).$$

If (3) is true, the first term in the inequality $\rightarrow 0$ as $T \rightarrow \infty$, and therefore so does the second. (5) is then true, and this, as shown above, implies (1). Next,

$$\begin{aligned} T \int_0^\infty x^{k-1} \sin^2(x/T) dG(x) &= T \int_0^T + T \int_T^\infty = I_1 + I_2; \\ I_1 &= T^{-1} \int_0^T x^{k+1} \left(\frac{\sin(x/T)}{x/T} \right)^2 dG(x); \\ \sin^2 1 \cdot T^{-1} \int_0^T x^{k+1} dG(x) &\leq I_1 \leq T^{-1} \int_0^T x^{k+1} dG(x). \end{aligned}$$

Hence $I_1 \rightarrow 0$ as $T \rightarrow \infty$ if and only if (2) is true. Thus (4) implies (2) which implies (1). Also

$$I_2 \leq T \int_T^\infty x^{k-1} dG(x),$$

and so $\rightarrow 0$ as $T \rightarrow \infty$ if (3) is true. Thus (1), which implies (2) and (3), implies (4).

LEMMA 2. When the statements 1-4 of Lemma 1 are true,

$$T \int_0^\infty x^{k-1} \sin(x/T) dG(x) - \int_0^T x^k dG(x) \rightarrow 0 \text{ as } T \rightarrow \infty.$$

This function of T is equal to

$$T \int_T^\infty x^{k-1} \sin(x/T) dG(x) - \int_0^T x^k \left(1 - \frac{\sin(x/T)}{x/T}\right) dG(x),$$

which has a modulus not greater than

$$T \int_T^\infty x^{k-1} dG(x) + \int_0^T x^k \cdot x^2/6T^2 \cdot dG(x) \leq T \int_T^\infty x^{k-1} dG(x) + \frac{1}{6}T^{-1} \int_0^T x^{k+1} dG(x).$$

This $\rightarrow 0$ as $T \rightarrow \infty$ because of (3) and (2).

4. Proof of theorem. If $\phi_0(t)$, $\phi_1(t)$ are the real and imaginary parts of $\phi(t)$,

$$\phi(t) = \phi_0(t) + i\phi_1(t),$$

$$\phi_0(t) = \int_{-\infty}^{\infty} \cos tx dF(x), \quad \phi_1(t) = \int_{-\infty}^{\infty} \sin tx dF(x).$$

$\phi_0(t)$ is an even function of t , and $\phi_1(t)$ is an odd function of t . A derivative of $\phi_0(t)$ of odd order which exists at $t = 0$ must be zero there, and the same is true of an even derivative of $\phi_1(t)$.

Let k be an odd positive integer, and suppose that $\phi^{(k)}(0)$ exists. It follows from the last paragraph that

$$\phi^{(k)}(0) = i\phi_1^{(k)}(0),$$

and so has real part zero. $\phi^{(k-1)}(0)$ must exist and be finite. As $k-1$ is even, this means that μ_{k-1} is finite [1]. Therefore $\phi^{(k-1)}(t)$ exists and is finite for all real t , and

$$\begin{aligned} \phi^{(k-1)}(t) &= i^{k-1} \int_{-\infty}^{\infty} x^{k-1} e^{itz} dF(x), \\ \frac{\phi^{(k-1)}(t) - \phi^{(k-1)}(0)}{t} &= i^{k-1} \int_{-\infty}^{\infty} x^{k-1} \frac{e^{itz} - 1}{t} dF(x) \\ (6) \quad &= -i^{k-1} \int_{-\infty}^{\infty} x^{k-1} \frac{\sin^2(\frac{1}{2}tx)}{\frac{1}{2}t} dF(x) \\ &\quad + i^k \int_{-\infty}^{\infty} x^{k-1} \frac{\sin tx}{t} dF(x). \end{aligned}$$

Put $G(x) = 1 - F(-x)$. This is a non-decreasing function of x . We may write

$$(7) \quad \begin{aligned} \frac{\phi^{(k-1)}(t) - \phi^{(k-1)}(0)}{t} &= -i^{k-1} \int_0^\infty x^{k-1} \frac{\sin^2(\frac{1}{2}tx)}{\frac{1}{2}t} d\{F(x) + G(x)\} \\ &+ i^k \left[\int_0^\infty x^{k-1} \frac{\sin tx}{t} dF(x) - \int_0^{1/t} x^k dF(x) \right] \\ &- i^k \left[\int_0^\infty x^{k-1} \frac{\sin tx}{t} dG(x) - \int_0^{1/t} x^k dG(x) \right] + i^k \int_{-1/t}^{1/t} x^k dF(x). \end{aligned}$$

Because $\phi^{(k)}(0)$ is purely imaginary, when $t \rightarrow 0$ the coefficient of i^{k-1} must $\rightarrow 0$. Hence $F(x)$ and $G(x)$ both satisfy (4) of Lemma 1 (with $T = 2/t$). Therefore they satisfy (1), i.e.,

$$T^k \{F(\infty) + G(\infty) - F(T) - G(T)\} \rightarrow 0 \text{ as } T \rightarrow \infty,$$

$$T^k \{1 - F(T) + F(-T)\} \rightarrow 0 \text{ as } T \rightarrow \infty,$$

which is equivalent to condition (i).

By Lemma 2 (with $T = 1/t$), the second and third terms on the right-hand side of (7) both $\rightarrow 0$ as $t \rightarrow 0$, and therefore

$$i^k \lim_{t \rightarrow 0} \int_{-1/t}^{1/t} x^k dF(x) = \phi^{(k)}(0).$$

Condition (ii) is thus necessary.

To prove that conditions (i) and (ii) are sufficient, suppose them satisfied. $F(x)$ and $G(x)$ satisfy (1) of Lemma 1 and therefore (3) also. Hence

$$\int_0^\infty x^{k-1} dF(x) \quad \text{and} \quad \int_0^\infty x^{k-1} dG(x)$$

are both finite, and

$$\mu_{k-1} = \int_0^\infty x^{k-1} d\{F(x) + G(x)\}$$

is finite. (6) is then true, and therefore (7). When $t \rightarrow 0$, the first and second terms on the right-hand side of (7) both $\rightarrow 0$, and the third term \rightarrow a limit. Thus $\phi^{(k)}(0)$ exists, and

$$\phi^{(k)}(0) = i^k \lim_{T \rightarrow \infty} \int_{-T}^T x^k dF(x).$$

REFERENCES

- [1] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 90.
- [2] A. ZYGMUND, "A Remark on characteristic functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 272-276.

N-DIMENSIONAL DISTRIBUTIONS CONTAINING A NORMAL COMPONENT¹

BY CHARLES STANDISH

Cornell University

In this paper we obtain necessary and sufficient conditions for an n -dimensional distribution function $F(x_1, \dots, x_n)$ to contain as a factor the distribution function of n independent normal random variables having common mean zero and variance 1. That is we obtain conditions for $F(x_1, \dots, x_n)$ to be of the form

$$(1) \quad F(x_1, \dots, x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} G(x_1 - u_1, \dots, x_n - u_n) dP(u_1, \dots, u_n),$$

where $P(u_1, \dots, u_n)$ is a distribution function and

$$G(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{\pi}}\right)^n \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \exp[-(u_1^2 + \dots + u_n^2)] du_1 \dots du_n.$$

If we denote $\partial^n / \partial x_1 \dots \partial x_n F(x_1, \dots, x_n)$ by $f(x_1, \dots, x_n)$, the problem becomes that of representing $f(x_1, \dots, x_n)$ in the form

$$(2) \quad f(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{\pi}}\right)^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\{-(x_1 - u_1)^2 + \dots + (x_n - u_n)^2\} dP(u_1, \dots, u_n).$$

The one-dimensional case has been treated by Pollard [1] employing properties of the heat equation. We use a different approach to prove the following

THEOREM. $f(x_1, \dots, x_n)$ is representable in the form (2) with $P(u_1, \dots, u_n)$ a distribution function if and only if

$$(i) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

(ii) $f(x_1, \dots, x_n)$ is bounded and has mixed partial derivatives of all orders satisfying

$$\left| \frac{\partial^{k_1} \dots \partial^{k_n}}{\partial x_1^{k_1} \dots \partial x_n^{k_n}} f(x_1, \dots, x_n) \right| \leq A^n 2 \frac{k_1 + \dots + k_n}{n} \sqrt{k_1! \dots k_n!},$$

$$k_1, \dots, k_n = 1, 2, \dots$$

$$(iii) \quad \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} \frac{(-1)^{k_1+\dots+k_n} t_1^{k_1} \dots t_n^{k_n}}{4^{k_1+\dots+k_n} k_1! \dots k_n!} \frac{\partial^{k_1} \dots \partial^{k_n}}{\partial x_1^{k_1} \dots \partial x_n^{k_n}} f(x_1, \dots, x_n) \geq 0,$$

$$|t_1| < 1, \dots, |t_n| < 1.$$

Received September 28, 1955.

¹ The research of the author was supported in part by the United States Air Force under Contract No. AF18(600)-685 monitored by the Office of Scientific Research.

PROOF. We carry out the proof for $n = 2$, the proof for $n > 2$ proceeding in exactly the same fashion. The necessity of (i) is obvious. As for (ii) we have

$$\left| \frac{\partial^{k_1}}{\partial x_1^{k_1}} \frac{\partial^{k_2}}{\partial x_2^{k_2}} f(x_1, x_2) \right| \leq \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |H_{k_1}(x_1 - u_1) H_{k_2}(x_2 - u_2)| \cdot \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dP(u_1, u_2),$$

where $H_k(x)$ is the k th Hermite polynomial which satisfies

$$(3) \quad |H_k(x)| \leq A 2^{k/2} \sqrt{k!} \exp \frac{x^2}{2}$$

([2], p. 236). Hence the integral above is majorized by

$$A^2 2^{\frac{k_1 + k_2}{2}} \sqrt{k_1! k_2!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -\frac{1}{2}[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dP(u_1, u_2),$$

which is $\leq A^2 2^{\frac{k_1 + k_2}{2}} \sqrt{k_1! k_2!}$. To establish the necessity of (iii) we observe

that we have formally

$$\begin{aligned} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \frac{(-1)^{k_1+k_2} t_1^{k_1} t_2^{k_2}}{4^{k_1+k_2} k_1! k_2!} \frac{\partial^{2k_1}}{\partial x_1^{2k_1}} \frac{\partial^{2k_2}}{\partial x_2^{2k_2}} f(x_1, x_2) \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \frac{(-1)^{k_1+k_2} t_1^{k_1} t_2^{k_2}}{4^{k_1+k_2} k_1! k_2!} H_{2k_1}(x_1 - u_1) H_{2k_2}(x_2 - u_2) \\ \times \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dP(u_1, u_2). \end{aligned}$$

From (3) it is seen that the double series in the integrand converges if all terms are replaced by their absolute values provided $|t_1| < 1$, $|t_2| < 1$, and the integral may be written as

$$(4) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\sum_{k_1=0}^{\infty} \frac{(-1)^{k_1} t_1^{k_1}}{4^{k_1} k_1!} H_{2k_1}(x_1 - u_1) \right] \left[\sum_{k_2=0}^{\infty} \frac{(-1)^{k_2} t_2^{k_2}}{4^{k_2} k_2!} H_{2k_2}(x_2 - u_2) \right] \\ \times \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dP(u_1, u_2),$$

but

$$\sum_{k_1=0}^{\infty} \frac{(-1)^{k_1} t_1^{k_1}}{4^{k_1} k_1!} H_{2k_1}(x) = \frac{1}{\sqrt{1-t_1}} \exp \left(-\frac{x^2 t_1}{1-t_1} \right)$$

([1], p. 580), and (4) becomes

$$(5) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{(1-t_1)(1-t_2)}} \exp - \left\{ \left[\frac{(x_1 - u_1)^2}{1-t_1} + \frac{(x_2 - u_2)^2}{1-t_2} \right] \right\} dP(u_1, u_2),$$

which for fixed t_1 and t_2 is \leq constant $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |dP(u_1, u_2)|$. This justifies the formal manipulations above and (5) is clearly non-negative establishing the necessity of (iii). For the sufficiency we need a couple of lemmas.

LEMMA 1. Denoting the left-hand side of (iii) by $T_{t_1, t_2} f(x_1, x_2)$ we have for functions $f(x_1, x_2)$ satisfying (ii)

$$\lim_{\substack{t_1 \rightarrow 1 \\ t_2 \rightarrow 1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} T_{t_1, t_2} f(u_1, u_2) du_1 du_2 = f(x_1, x_2).$$

PROOF. The estimates furnished by (ii) enable us to write

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} \frac{\partial^{2k_1}}{\partial u_1^{2k_1}} \frac{\partial^{2k_2}}{\partial u_2^{2k_2}} f(u_1, u_2) du_1 du_2 \\ &= \int_{-\infty}^{\infty} \exp [-(x_2 - u_2)^2] du_2 \int_{-\infty}^{\infty} \exp [-(x_1 - u_1)^2] \frac{\partial^{2k_1}}{\partial u_1^{2k_1}} \frac{\partial^{2k_2}}{\partial u_2^{2k_2}} f(u_1, u_2) du_1, \end{aligned}$$

and upon integrating the inner integral $2k_1$ times by parts we have

$$\begin{aligned} & \int_{-\infty}^{\infty} \exp [-(x_2 - u_2)^2] du_2 \int_{-\infty}^{\infty} \frac{\partial^{2k_1}}{\partial u_1^{2k_1}} \exp [-(x_1 - u_1)^2] \frac{\partial^{2k_2}}{\partial u_2^{2k_2}} f(u_1, u_2) du_1 \\ &= \int_{-\infty}^{\infty} \frac{\partial^{2k_1}}{\partial u_1^{2k_1}} \exp [-(x_1 - u_1)^2] du_1 \int_{-\infty}^{\infty} \exp [-(x_2 - u_2)^2] \frac{\partial^{2k_2}}{\partial u_2^{2k_2}} f(u_1, u_2) du_2. \end{aligned}$$

We integrate $2k_2$ more times by parts and obtain finally

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^{2k_1}}{\partial u_1^{2k_1}} \exp [-(x_1 - u_1)^2] \frac{\partial^{2k_2}}{\partial u_2^{2k_2}} \exp [-(x_2 - u_2)^2] f(u_1, u_2) du_1 du_2.$$

Thus

$$\begin{aligned} & \lim_{\substack{t_1 \rightarrow 1 \\ t_2 \rightarrow 1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} T_{t_1, t_2} f(u_1, u_2) du_1 du_2 \\ &= \lim_{\substack{t_1 \rightarrow 1 \\ t_2 \rightarrow 1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\sum_{k_1=0}^{\infty} \frac{(-1)^{k_1} t_1^{k_1}}{4^{k_1} k_1!} H_{2k_1}(x_1 - u_1) \left[\sum_{k_2=0}^{\infty} \frac{(-1)^{k_2} t_2^{k_2}}{4^{k_2} k_2!} H_{2k_2}(x_2 - u_2) \right] \right. \\ & \quad \times \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} f(u_1, u_2) du_1 du_2. \end{aligned}$$

By (4) and (5) this becomes

$$\begin{aligned} & \left(\lim_{t_1 \rightarrow 1} \frac{1}{\sqrt{1-t_1}} \int_{-\infty}^{\infty} \exp \left[-\frac{(x_1 - u_1)^2}{1-t_1} \right] du_1 \right) \\ & \quad \times \left(\lim_{t_2 \rightarrow 1} \frac{1}{\sqrt{1-t_2}} \int_{-\infty}^{\infty} \exp \left[-\frac{(x_2 - u_2)^2}{1-t_2} \right] f(u_1, u_2) du_2 \right) = f(x_1, x_2). \end{aligned}$$

LEMMA 2.

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_{t_1, t_2} f(u_1, u_2) du_1 du_2 = 1, \quad |t_1| < 1, |t_2| < 1.$$

PROOF.

$$\begin{aligned}
 & \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_{t_1, t_2} f(u_1, u_2) du_1 du_2 \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1 dx_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} T_{t_1, t_2} f(u_1, u_2) du_1 du_2 \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1 dx_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{(1-t_1)(1-t_2)}} \exp \left\{ -\left[\frac{(x_1 - u_1)^2}{1-t_1} + \frac{(x_2 - u_2)^2}{1-t_2} \right] \right\} \\
 &\quad \times f(u_1, u_2) du_1 du_2 \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u_1, u_2) du_1 du_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{(1-t_1)(1-t_2)}} \\
 &\quad \times \exp \left\{ -\left[\frac{(x_1 - u_1)^2}{1-t_1} + \frac{(x_2 - u_2)^2}{1-t_2} \right] \right\} dx_1 dx_2 = 1.
 \end{aligned}$$

By the above lemma and (iii) the family of functions

$$P_{t_1, t_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} T_{t_1, t_2} f(u_1, u_2) du_1 du_2$$

is monotone in the sense of Bochner ([3], p. 383) and uniformly bounded; hence there exist sequences $\{t_{1n}\}$ $\{t_{2n}\}$ such that $t_{1n} \rightarrow 1$, $t_{2n} \rightarrow 1$ and a function $P(x_1, x_2)$ monotone and bounded such that

$$\lim_{n \rightarrow \infty} P_{t_{1n}, t_{2n}}(x_1, x_2) = P(x_1, x_2)$$

([3], p. 389-390). By Lemma 1,

$$f(x_1, x_2) = \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dP_{t_{1n}, t_{2n}}(u_1, u_2).$$

By the formula for integration by parts in two dimensions [4] the above integral becomes

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{t_{1n}, t_{2n}}(u_1, u_2) \frac{\partial^2}{\partial u_1 \partial u_2} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} du_1 du_2,$$

and integrating by parts again

$$f(x_1, x_2) = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dP(u_1, u_2).$$

To complete the proof that $P(u_1, u_2)$ is a distribution function we observe that by condition (1)

$$\begin{aligned}
 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dP(u_1, u_2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ -[(x_1 - u_1)^2 + (x_2 - u_2)^2] \} dx_1 dx_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dP(u_1, u_2).
 \end{aligned}$$

REFERENCES

- [1] H. POLLARD, "Distribution functions containing a Gaussian factor," *Proc. Am. Math. Soc.*, Vol. 4 (1953), pp. 578-582.
- [2] G. SZEGO, *Orthogonal Polynomials*, American Mathematical Society New York, 1939.
- [3] S. BOCHNER, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Math. Annalen*, Vol. 108 (1933).
- [4] W. H. YOUNG, "On multiple integration by parts and the second theorem of the mean," *Proc. London Math. Soc.*, Vol. 16 (1917), p. 276.

A CERTAIN CLASS OF TESTS OF FIT¹

BY LIONEL WEISS

University of Oregon

1. Summary and introduction. Suppose X_1, X_2, \dots, X_n are known to be independently and identically distributed, each with the density function $f(x)$, with $\int_0^1 f(x) dx = 1$. Let $Y_1 \leq Y_2 \leq \dots \leq Y_n$ be the ordered values of X_1, X_2, \dots, X_n , and define $W_1 = Y_1, W_2 = Y_2 - Y_1, \dots, W_n = Y_n - Y_{n-1}$, and $W_{n+1} = 1 - Y_n$, so that $W_1 + \dots + W_{n+1} = 1$. Finally, define Z_1, \dots, Z_{n+1} as the ordered values of W_1, \dots, W_{n+1} , so that $0 \leq Z_1 \leq Z_2 \leq \dots \leq Z_{n+1}$, with $Z_1 + \dots + Z_{n+1} = 1$. We are going to test the hypothesis that $f(x) = 1$ for $0 < x < 1$, and we are going to consider only tests based on Z_1, Z_2, \dots, Z_n . The intuitive justification for this is that, roughly speaking, deviations from the hypothesis on any part of the unit interval are treated alike. Several authors have discussed tests based on Z_1, \dots, Z_n . (See references [1], [2], [3].)

If u is a number greater than unity, it is shown that the test of the form "reject the hypothesis if $Z_1^u + \dots + Z_{n+1}^u > K$ " is consistent against a very wide class of alternatives. When $u = 2$, the resulting test has some desirable properties with respect to alternatives with linear density functions.

2. The distribution of Z_1, \dots, Z_n . It is easily seen that $P[Z_i = Z_j \text{ for any } i \neq j]$ is equal to zero. We want to find the joint density function $h(z_1, \dots, z_n)$ of Z_1, \dots, Z_n . The joint density function of W_1, \dots, W_n is equal to $n! f(w_1)f(w_1 + w_2) \dots f(w_1 + w_2 + \dots + w_n)$ in the region $w_i \geq 0, w_1 + \dots + w_n \leq 1$, and is equal to zero elsewhere. Let $\{j(1), j(2), \dots, j(n+1)\}$ be any permutation of the first $n+1$ integers, and let \sum_p denote summation over all the $(n+1)!$ permutations. Given any set of numbers $0 < z_1 < z_2 < \dots < z_n < 1 - (z_1 + \dots + z_n)$, we denote by $Q[j(1), j(2), \dots, j(n+1)]$ the conditional probability that $W_i = z_{j(i)}$ for $i = 1, \dots, n+1$, given that $Z_i = z_i$ for $i = 1, \dots, n+1$. It is understood that if $j(i) = n+1$, then $z_{j(i)} = 1 -$

Received August 22, 1955; revised June 20, 1956.

¹ Research under contract with the Office of Naval Research.

$(z_1 + \dots + z_n)$. For each set of values z_1, \dots, z_n for which $h(z_1, \dots, z_n)$ is positive, we have

$$Q[j(1), \dots, j(n+1)] = \frac{n! f(z_{j(1)}) f(z_{j(1)} + z_{j(2)}) \dots f(z_{j(1)} + \dots + z_{j(n)})}{h(z_1, \dots, z_n)}.$$

Since $\sum_p Q[j(1), \dots, j(n+1)] = 1$, for each set of values z_1, \dots, z_n for which $h(z_1, \dots, z_n)$ is positive, we have $h(z_1, \dots, z_n) = n! \sum_p f(z_{j(1)}) f(z_{j(1)} + z_{j(2)}) \dots f(z_{j(1)} + \dots + z_{j(n)})$. Now let D be the region in (z_1, \dots, z_n) -space where the following three conditions are satisfied:

- (1) $0 \leq z_1 \leq z_2 \leq \dots \leq z_n \leq 1 - (z_1 + \dots + z_n)$,
- (2) $h(z_1, \dots, z_n) = 0$,
- (3) $n! \sum_p f(z_{j(1)}) f(z_{j(1)} + z_{j(2)}) \dots f(z_{j(1)} + \dots + z_{j(n)}) > 0$.

Then D must be of measure zero. For if D is of positive measure, then

$$\int \dots \int_D n! \sum_p f(z_{j(1)}) f(z_{j(1)} + z_{j(2)}) \dots f(z_{j(1)} + \dots + z_{j(n)}) dz_1 \dots dz_n > 0,$$

which implies that $P[(W_1, \dots, W_n) \text{ in } D] > 0$, which in turn implies that $P[(Z_1, \dots, Z_n) \text{ in } D] > 0$, which is a contradiction. Therefore we have shown that if $0 \leq z_1 \leq z_2 \leq \dots \leq z_n \leq 1 - (z_1 + \dots + z_n)$, then

$$(2.1) \quad h(z_1, \dots, z_n) = n! \sum_p f(z_{j(1)}) f(z_{j(1)} + z_{j(2)}) \dots f(z_{j(1)} + \dots + z_{j(n)}),$$

while $h(z_1, \dots, z_n)$ is zero for other values of z_1, \dots, z_n . We note that when $f(x) = 1$, the right-hand side of (2.1) is equal to $n!(n+1)!$.

3. Properties of the power of tests based on Z_1, \dots, Z_n . Let $r(x)$ be a given bounded measurable function of x satisfying the conditions

$$\int_0^1 r(x) dx = 0, \quad \int_0^1 r^2(x) dx > 0.$$

Then for δ small enough in absolute value, $1 + \delta r(x)$ is a density function on $(0, 1)$. For any given measurable region R in (z_1, \dots, z_n) -space, we denote by $M(R, \delta)$ the probability that (Z_1, \dots, Z_n) will fall in R , assuming the density of the original observations is equal to $1 + \delta r(x)$. In what follows, we shall always assume that R is a subset of the region

$$0 \leq z_1 \leq \dots \leq z_n \leq 1 - (z_1 + \dots + z_n).$$

For any given region R , we have

$$(3.1) \quad \left. \frac{dM(R, \delta)}{d\delta} \right|_{\delta=0} = n! \int \dots \int_R \sum_p \sum_{k=1}^n r(z_{j(1)} + \dots + z_{j(k)}) dz_1 \dots dz_n,$$

$$(3.2) \quad \left. \frac{d^2 M(R, \delta)}{d\delta^2} \right|_{\delta=0} = 2n! \int \dots \int_R \sum_p \sum_{1 \leq k < l \leq n} r(z_{j(1)} + \dots + z_{j(k)}) \cdot r(z_{j(1)} + \dots + z_{j(l)}) dz_1 \dots dz_n.$$

Equations (3.1) and (3.2) follow easily when $f(x)$ in (2.1) is replaced by $1 + \delta r(x)$, and the result is expressed as a polynomial in δ .

4. The case of linear $r(x)$. The integrands in (3.1) and (3.2) are complicated for many functions $r(x)$. However, when $r(x) = x - \frac{1}{2}$, we have (remembering $z_{n+1} = 1 - (z_1 + \dots + z_n)$) that the integrand in (3.1) is identically equal to zero, while the integrand in (3.2) is equal to

$$(4.1) \quad \frac{(n-1)(n+2)!}{24} (z_1^2 + \dots + z_{n+1}^2) - \frac{(n-1)(n+1)!}{12}.$$

Suppose we are testing the hypothesis that $f(x) = 1$ against alternatives of the form $f(x) = 1 + \delta(x - \frac{1}{2})$, with given level of significance α . We are going to consider only tests based on Z_1, \dots, Z_n , so that our critical region will be a region in (Z_1, \dots, Z_n) -space. We want to find the critical region R satisfying the following three conditions:

- (1) $M(R, 0) = \alpha$,
- (2) $\left. \frac{dM(R, \delta)}{d\delta} \right|_{\delta=0} = 0$,
- (3) $\left. \frac{d^2 M(R, \delta)}{d\delta^2} \right|_{\delta=0}$ is a maximum.

In the terminology of [6], this region R would be called an "unbiased critical region of type A" for testing the hypothesis that $\delta = 0$. We know that in the present case, condition (2) is automatically satisfied by any region R , since the integrand in (3.1) is identically zero. But then a very simple application of the Neyman-Pearson lemma shows that the desired region R is given by

$$\frac{2n! \{ (n-1)(n+2)(z_1^2 + \dots + z_{n+1}^2)/24 - (n-1)(n+1)/12 \}}{n!(n+1)!} \geq K(\alpha),$$

where $K(\alpha)$ is a properly chosen constant. Equivalently, R is given by

$$z_1^2 + \dots + z_{n+1}^2 \geq k(\alpha),$$

where $k(\alpha)$ is a properly chosen constant.

5. Consistency of the proposed test. In this section we prove that the test described in Section 4 is one of a class of tests, any one of which is consistent against a wide class of alternatives. First we need some lemmas.

LEMMA 1. If $g(x)$, the common density of X_1, \dots, X_n , has at most a finite number of discontinuities, and if $R_n(t)$ denotes the proportion of the values

$$Z_1, \dots, Z_{n+1}$$

which are not greater than $t/(n+1)$, while $S(t)$ denotes $1 - \int_0^1 g(x) \exp \{-t g(x)\} dx$, and $V(n)$ denotes $\sup_{t \geq 0} |R_n(t) - S(t)|$, then $V(n)$ converges to zero with probability one as n increases.

PROOF. This is proved in [4].

Now we introduce the following notation. Let u be any positive number, while Y_n shall denote $\Gamma(n+u+1)/\Gamma(n+2)[Z_1^u + \dots + Z_{n+1}^u]$. Let $g(x)$ denote the common density of X_1, \dots, X_n , and define $J(g; u)$ as

$$\Gamma(u+1) \int_{\{x: g(x) > 0\}} [g(x)]^{1-u} dx.$$

$J(g; u)$ may fail to exist (that is, be infinite).

LEMMA 2. If $J(g; u)$ is finite, then given any positive numbers ϵ, δ , there is a positive integer $N(\epsilon, \delta)$ such that

$$P[Y_n > J(g; u) - \epsilon \text{ simultaneously for all } n > N(\epsilon, \delta)] > 1 - \delta.$$

If $J(g; u)$ fails to exist, then given any positive numbers B, δ , there is a positive integer $M(B, \delta)$ such that

$$P[Y_n > B \text{ simultaneously for all } n > M(B, \delta)] > 1 - \delta.$$

PROOF. In the notation of Lemma 1, we have

$$Z_1^u + \dots + Z_{n+1}^u = (n+1)^{1-u} \int_0^\infty t^u dR_n(t),$$

and therefore $Y_n = \Gamma(n+u+1)/\Gamma(n+2)(n+1)^{1-u} \int_0^\infty t^u dR_n(t)$. Now we choose any positive number T and hold it fixed until further notice. We have $Y_n \geq \Gamma(n+u+1)/\Gamma(n+2)(n+1)^{1-u} \int_0^T t^u dR_n(t)$. As n increases, the coefficient of the integral in this last expression approaches unity, and from now on we shall treat it as unity, and it will be seen that this does not affect our conclusion. We have $\int_0^T t^u dR_n(t) = T^u R_n(T) - u \int_0^T t^{u-1} R_n(t) dt$, and by Lemma 1, the expression on the right of this equality approaches the following with probability one:

$$T^u \left[1 - \int_0^1 g(x) \exp \{-Tg(x)\} dx \right] - u \int_0^T t^{u-1} dt \\ + u \int_0^1 g(x) \left\{ \int_0^T t^{u-1} \exp [-tg(x)] dt \right\} dx,$$

which equals

$$- \int_0^1 T^u g(x) \exp [-Tg(x)] dx + u \int_{\{x: g(x) > 0\}} [g(x)]^{1-u} \int_0^{Tg(x)} r^{u-1} e^{-r} dr dx.$$

But by taking T large enough, this last expression can be made arbitrarily close to $J(g; u)$ if it exists, or it can be made arbitrarily large if $J(g; u)$ fails to exist. This proves Lemma 2.

LEMMA 3. If the common density of X_1, \dots, X_n is uniform on $(0, 1)$, then Y_n converges stochastically to $\Gamma(u+1)$ as n increases.

PROOF. This is proved directly from the discussion on page 245 of [5].

LEMMA 4. If $u > 1$, and if $g(x)$ is positive almost everywhere on $(0, 1)$ and differs from unity on a subset of $(0, 1)$ of positive measure, then $J(g; u) > \Gamma(u + 1)$.

PROOF. For convenience, we omit the limits of integration, which are always zero and one throughout this proof. Hölder's inequality states that if $p > 0$, $q > 0$, and $p + q = pq$, then

$$\left| \int r(x)s(x) dx \right| \leq \left(\int |r(x)|^p dx \right)^{1/p} \left(\int |s(x)|^q dx \right)^{1/q},$$

with equality holding if and only if $|r(x)|^p = K |s(x)|^q$ almost everywhere, where K is a constant, and either $r(x)s(x) \geq 0$ almost everywhere or $r(x)s(x) \leq 0$ almost everywhere. Applying this inequality with $r(x) = [g(x)]^{(u-1)/u}$, $s(x) = [g(x)]^{-1/u}$, $p = u/(u-1)$, and $q = u$, the lemma follows immediately.

THEOREM. Suppose it is known that X_1, X_2, \dots are independently and identically distributed, and it is desired to test the hypothesis that the common distribution is the uniform distribution over $(0, 1)$. For a given level of significance, α ($0 < \alpha < 1$), a given number $u > 1$, and a given positive integer n , let $T(\alpha; n; u)$ denote the test of the hypothesis described as follows: Reject the hypothesis if and only if at least one of the following occurs:

- (1) At least one of the values X_1, \dots, X_n falls outside the open interval $(0, 1)$,
- (2) $X_i = X_j$ for some integers i, j with $1 \leq i < j \leq n$,
- (3) $Z_1^n + Z_2^n + \dots + Z_{n+1}^n \geq K(\alpha; n; u)$,

where $K(\alpha; n; u)$ is a constant chosen to give the proper level of significance. Then the sequence of tests $\{T(\alpha; n; u), T(\alpha; n+1; u), \dots\}$ is consistent against any alternative common distribution function $G(x)$ with at least one of the following properties:

- (1') $G(0) > 0$,
- (2') $G(1) < 1$,
- (3') $G(x)$ has at least one positive saltus,
- (4') $G(0) = 0, G(1) = 1, G(x)$ is absolutely continuous with derivative $g(x)$, and $g(x)$ differs from unity on a subset of $(0, 1)$ of positive measure and has at most a finite number of discontinuities, and a finite number of oscillations.

PROOF. If $G(x)$ has property (1') or (2'), specification (1) of $T(\alpha; n; u)$ proves consistency. If $G(x)$ has property (3'), specification (2) of $T(\alpha; n; u)$ proves consistency. If property (4') is possessed by $G(x)$, we distinguish two cases, according to whether or not $g(x)$ is positive almost everywhere on $(0, 1)$.

CASE 1: $g(x)$ is positive almost everywhere on $(0, 1)$. We may express specification (3) of the test in terms of Y_n defined above. For large n , Lemma 3 tells us that specification (3) of the test is essentially $Y_n > \Gamma(u + 1)$. But Lemmas 2 and 4 guarantee that under $G(x)$ the probability is high that Y_n will be greater than $\Gamma(u + 1)$ if n is large.

CASE 2: $g(x)$ is zero on a subset of $(0, 1)$ of positive measure. Since $g(x)$ has at most a finite number of discontinuities, a point w in the interior of $(0, 1)$ can be found such that $g(x)$ is continuous in a neighborhood of w , $g(w) = 0$, and any neighborhood of w contains a set of positive measure on which $g(x) = 0$. Since

$g(x)$ has a finite number of oscillations, this implies that there is an interval of positive length Δ in the interior of $(0, 1)$ on which $g(x) = 0$. But then the largest of the values Z_1, \dots, Z_{n+1} is certainly no smaller than Δ ; therefore Y_n is certainly no smaller than $\Delta^n \Gamma(n+u+1)/\Gamma(n+2)$, and this last expression approaches infinity as n increases.

REFERENCES

- [1] B. F. KIMBALL, "Some basic theorems for developing tests of fit for the case of nonparametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540-548.
- [2] B. SHERMAN, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339-361.
- [3] M. GREENWOOD, "The statistical study of infectious diseases," *J. Roy. Stat. Soc.*, Vol. 109 (1946), pp. 85-110.
- [4] L. WEISS, "The stochastic convergence of a function of sample successive differences," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 532-536.
- [5] D. A. DARLING, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239-253.
- [6] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1-37.

ON THE PROBABILITY OF LARGE DEVIATIONS FOR SUMS OF BOUNDED CHANCE VARIABLES

BY HARRY WEINGARTEN

Bureau of Ships, Navy Department

1. Summary. The following theorems are proved.

THEOREM 1. If x_1, x_2, \dots satisfy $-1 \leq x_n \leq a$, $a \leq 1$ and

$$E(x_n | x_1, \dots, x_{n-1}) \leq -u \max(|x_n| | x_1, \dots, x_{n-1}),$$

$0 < u < 1$, then for any positive t ,

$$\Pr \{x_1 + \dots + x_n \geq t \text{ for some } n\} \leq \theta^t,$$

where θ is the positive root (other than $\theta = 1$) of

$$(1) \quad \frac{a+u}{a+1} \theta^{a+1} - \theta^a + \frac{1-u}{a+1} = 0.$$

This choice of θ is the best possible.

THEOREM 2. If x_1, x_2, \dots satisfy $|x_n| \leq 1$ and $E(x_n | x_1, \dots, x_{n-1}) = 0$, then for all $N > 0$,

$$\Pr \left\{ \left| \frac{x_1 + \dots + x_n}{n} \right| \geq \epsilon \text{ for some } n \geq N \right\} \leq 2e^{-N},$$

Received November 4, 1955. (Revised January 3, 1956).

where $\varphi = (1 + \epsilon)^{-(1+\epsilon)/2}(1 - \epsilon)^{-(1-\epsilon)/2}$. This choice of φ is, for every ϵ between 0 and 1, the best possible.

Both results are improvements of results of Blackwell [1], and the methods of proof are somewhat similar.

2. Proofs of the Theorems. Since the methods for Theorems 1 and 2 are similar to those in [1], we merely indicate the main steps.

For Theorem 1, let $\Phi(N, t)$ be the least upper bound, over all sequences $\{x_n\}$ satisfying the hypotheses of Theorem 1, of the probability

$$\Pr \{x_1 + \cdots + x_k \geq t \text{ for some } k \leq N\};$$

in particular $\Phi(0, t)$ is 1 for $t \leq 0$ and 0 for $t > 0$. Then

$$\Phi(N + 1, t) = U\Phi(N, t),$$

where U is the transformation taking Borel-measurable functions of t into Borel-measurable functions of t , such that the value of Uf at t is

$$\sup_{x \in X} Ef(t - x),$$

where X consists of all chance variables satisfying $-1 \leq x \leq a$ and $EX \leq -u \max |X|$. Now if θ satisfies (1), then $Ug = g$, where $g = \theta^t$. Also, $f_1 \geq f_2$ for all t implies $Uf_1 \geq Uf_2$ for all t . Repeated application of this to $g \geq \Phi(0, t)$ yields $g \geq \Phi(N, t)$ for all t , and letting $N \rightarrow \infty$ completes the proof of Theorem 1. To see that this choice is the best possible consider the sequence x_1, x_2, \dots independent, with the distribution of each $x_n = a$, and -1 with probabilities $(1 - u)/(a + 1)$, $(a + u)/(a + 1)$ respectively. This sequence satisfies the hypotheses of Theorem 1, and it will be shown that

$$\Pr \{x_1 + \cdots + x_n \geq t \text{ for some } n\}^{1/n} \rightarrow \theta$$

as $t \rightarrow \infty$.

To do this we consider a game between two players with fortunes, stakes, etc., as follows:

Players.....	P_1	P_2
Fortunes.....	t	b
Stakes.....	$a \leq 1$	1
Probability of winning a game.....	$p = \frac{a+u}{a+1}$	$q = \frac{1-u}{a+1}$

The probability of the ruin of P_1 which we are interested in is easily seen to be the same as $\Pr \{x_1 + \cdots + x_n \geq t\}$ for the case of a sequence x_1, x_2, \dots independently and identically distributed with each $x_n = a$, -1 with probabilities $(1 - u)/(a + 1)$, $(a + u)/(a + 1)$ respectively.

Let us approximate a by some rational fraction r/s and then change the units in which the game is played. We will have

Players.....	P_1	P_2
Fortunes.....	st	sb
Stakes.....	r	s
Probability of winning a game.....	$p = \frac{a+u}{a+1}$	$q = \frac{1-u}{a+1}$

Using the results of [2], pages 144-146, we obtain

$$\theta_1^{st} \frac{\theta_1^{b-s+1} - 1}{\theta_1^{s(b-s+1)} - 1} \leq y_{st} \leq \theta_1^{st-r+1} \frac{\theta_1^{sb} - 1}{\theta_1^{s(b-s-r+1)} - 1},$$

where θ_1 is the root of $p\theta_1^{r+s} - \theta_1^r + q = 0$ and y_{st} is the probability of the ruin of P_1 when his fortune is st . If the fortune of P_2 becomes infinite, we have

$$\theta_1^{st} \leq y_{st} \leq \theta_1^{st-r+1}.$$

When we return to the original units of the game, we can state

$$[\theta_1^s]^t \leq y_t \leq [\theta_1^s]^{t-(r/s)+(1/s)}$$

or

$$\theta_2^t \leq y_t \leq \theta_2^{t-(r/s)+(1/s)},$$

where θ_2 is the root of $p\theta_2^{(r/s)+1} - \theta_2^{r/s} + q = 0$.

By choosing r and s large enough, we may come as close as we wish to a , and so we may finally write

$$\theta^t \leq y_t \leq \theta^{t-a}$$

where θ is the root of $p\theta^{a+1} - \theta^a + q = 0$. This is possible, since the probability of ruin in the game where the stakes are r and s is the general solution of the difference equation

$$y_x = py_{x+s} + qy_{x-r},$$

where y_x is the probability of ruin of P_1 when his fortune is x . Such solutions are known to be continuous functions of the stakes. That θ_2 , the root of $p\theta_2^{(r/s)+1} - \theta_2^{r/s} + q = 0$, approaches θ , the root of $p\theta^{a+1} - \theta^a + q = 0$, follows from the fact that the solution of a polynomial is a continuous function of the coefficients. From this we may obtain

$$[\theta^t]^{1/t} \leq \Pr \{x_1 + \dots + x_n \geq t\}^{1/t} \leq [\theta^{t-a}]^{1/t}$$

and so

$$\Pr \{x_1 + \dots + x_n \geq t\}^{1/t} \rightarrow \theta \quad \text{as} \quad t \rightarrow \infty.$$

As a matter of fact we note that θ is really a lower bound, since

$$\theta^{1-(a/t)} > \theta.$$

Hence θ is best possible.

For the proof of Theorem 2 we have:

$$\begin{aligned} & \Pr \left\{ \frac{x_1 + \cdots + x_n}{n} \geq \epsilon \text{ for some } n \geq N \right\} \\ & \leq \Pr \{x_1 - k\epsilon + \cdots + x_n - k\epsilon \geq N\epsilon(1-k) \text{ for some } n\} \\ & \leq [(1+\epsilon)^{-(1+\epsilon)/2} \cdot (1-\epsilon)^{-(1-\epsilon)/2}]^N, \end{aligned}$$

where the last inequality is obtained by applying Theorem 1 to the sequence

$$y_n = \frac{x_n - k\epsilon}{1 + k\epsilon}.$$

Here we have taken

$$a = \frac{1 - k\epsilon}{1 + k\epsilon} \quad \text{and} \quad u = \frac{k\epsilon}{1 + k\epsilon}$$

and found

$$\Pr \left\{ y_1 + \cdots + y_n \geq \frac{N\epsilon(1-k)}{1+k\epsilon} \right\} \leq \theta^{N\epsilon(1-k)/(1+k\epsilon)},$$

where θ is the root of

$$(2) \quad \frac{1}{2}\theta^{2/(1+k\epsilon)} - \theta^{(1-k\epsilon)/(1+k\epsilon)} + \frac{1}{2} = 0.$$

To find the smallest value of $\theta^{N\epsilon(1-k)/(1+k\epsilon)}$ we may proceed in the following manner: Beginning with (2), we write

$$(3) \quad \theta^{(1-k\epsilon)/(1+k\epsilon)}(\theta - 2) + 1 = 0,$$

and solving for $k\epsilon$, we find

$$(4) \quad k\epsilon = \frac{\log \theta + \log (2 - \theta)}{\log \theta - \log (2 - \theta)}.$$

Giving $k\epsilon$ the value from (4), we find that $R = \theta^{N\epsilon(1-k)/(1+k\epsilon)}$ gives

$$(5) \quad R^{1/N} = \theta^{-[(1-\epsilon) \log \theta + (1+\epsilon) \log (2-\theta)]/2 \log \theta}$$

If we take logarithms in both sides of (5) and simplify, we may rewrite (5) to obtain

$$(6) \quad R^{1/N} = \theta^{-(1-\epsilon)/2} (2 - \theta)^{-(1+\epsilon)/2},$$

and it is very easy to find the value of θ which makes R a minimum to be

$$(7) \quad \theta = 1 - \epsilon.$$

The same inequality holds for

$$\Pr \left\{ \frac{x_1 + \cdots + x_n}{n} \leq -\epsilon \text{ for } n \geq N \right\},$$

and hence Theorem 2 is proved.

To see that φ is the best possible, consider the case of the sequence $\{x_n\}$ independently distributed, each taking the values ± 1 with probabilities $\frac{1}{2}, \frac{1}{2}$. It follows from a result of Chernoff [3] that

$$\Pr \{x_1 + \cdots + x_n \geq n\epsilon\}^{1/n} \rightarrow \varphi \quad \text{as} \quad n \rightarrow \infty,$$

so that our φ is exact.

REFERENCES

- [1] D. BLACKWELL, "On optimal systems," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 394-397.
- [2] J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill, 1937.
- [3] H. CHERNOFF, "A measure of asymptotic efficiency for test of a hypothesis based on a sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493-507.

A REMARK ON THE ROOTS OF THE MAXIMUM LIKELIHOOD EQUATION

BY C. KRAFT AND L. LECAM¹

University of California, Berkeley

1. Introduction and summary. The statistical literature combines two types of investigations concerning the consistency of maximum likelihood (M.L.) estimates. A few of these, such as the most excellent paper of A. Wald [1], do prove directly the consistency of M.L. estimates. However, most investigators seem to have concentrated their efforts on proving the existence and consistency of suitably selected roots of the successive likelihood equations. Some authors, see [2], for example, add the supplementary remark that such consistent roots will eventually be unique in suitably small neighborhoods of the true value and will achieve a local maximum.

It is the purpose of the present note to point out by means of examples that this second mode of attack is not adequate. In the examples given below, the "usual regularity conditions" of Cramér [3] or Wald [4] are satisfied, but the M.L. estimates are not consistent. It should also be pointed out that the direct proofs of existence of roots, simple in the case of a unidimensional parameter, become unwieldy in more than one dimension. On the other hand, if one has proved the consistency of the M.L. estimates, the existence of roots follows trivially from the fact that when a differentiable function reaches its maximum in an open set, the derivatives vanish at that point.

2. Examples with independent identically distributed variables. The first example given below has the following characteristics:

Received September 21, 1955.

¹ This paper was prepared with the support of the Office of Ordnance Research, U. S. Army, under Contract DA-04-200-ORD-171.

(1) Cramér's conditions are satisfied and the condition of identifiability is satisfied.

(2) The likelihood equation has roots.

(3) The M.L. estimate does not exist, except maybe for sets of sample points of measure zero.

(4) There exist consistent estimates. (See Section 3, below.)

For every nonnegative integer k , let A_k be the open interval

$$A_k = (2k, 2k + 1).$$

Let $\Theta = \bigcup_{k=1}^{\infty} A_k$, and let $\{a_k\}$, $k = 0, 1, 2, \dots$, be an arbitrary ordering of the rationals of the interval $(1, 2)$. Define $\rho(\theta) = a_k$ if $\theta \in A_k$. For each $\theta \in \Theta$, let the vector (X, Y) have a normal distribution with $E(X | \theta) = \rho(\theta) \cos 2\pi\theta$ and $E(Y | \theta) = \rho(\theta) \sin 2\pi\theta$, and covariance matrix the identity. If $\{(X_i, Y_i)\}$, $i = 1, 2, \dots, n$, is a sequence of independent random vectors with the distribution of (X, Y) , the logarithm of the probability density of the first n observation is given by

$$-2 \log p_n = K_n + n[\bar{X}_n - \rho(\theta) \cos 2\pi\theta]^2 + n[\bar{Y}_n - \rho(\theta) \sin 2\pi\theta]^2,$$

where (\bar{X}_n, \bar{Y}_n) is the sample mean.

Defining $r_n > 0$ and φ_n , $0 \leq \varphi_n < 1$, by $\bar{X}_n = r_n \cos 2\pi\varphi_n$ and $\bar{Y}_n = r_n \sin 2\pi\varphi_n$, the above equation can also be written as

$$-2 \log p_n = n \log 2\pi + [r_n - \rho(\theta)]^2 + 2r_n\rho(\theta)[1 - \cos 2\pi(\theta - \varphi_n)].$$

Accordingly, the likelihood equation is $r_n \sin 2\pi(\theta - \varphi_n) = 0$. Therefore, all values of the form $\theta = \varphi_n + k/2$ which belong to Θ are solutions of the likelihood equation. However, if r_n is not rational, the M.L. estimate does not exist, since $\rho(\theta)$ can be chosen close to r_n but not equal to it.

One could define approximate maximum likelihood estimates as follows. Let $\{\epsilon_n\}$ be a sequence of positive numbers tending to zero. For each n , let S_n be the set of values of θ such that

$$\sup_{i \in \theta} p_n(x_1, x_2, \dots, x_n | i) \leq (1 + \epsilon_n)p_n(x_1, \dots, x_n | \theta).$$

Since every interval, however small, contains an infinity of rationals, for every $\epsilon_n > 0$, the set S_n will, in our example, have elements in common with an infinite number of the intervals A_k , and therefore the sequence $\{S_n\}$ cannot converge to a point.

One might object to the preceding example for two reasons. In the usual proofs of "consistency" of roots of the M.L. equation, it is assumed that the random variables are real-valued. However, this assumption is irrelevant to the proofs given, so that the bivariate character of the example is no detraction. It is, of course, possible to build analogous univariate examples.

Another feature to which objections can be raised is the nonexistence of the M.L. estimate. This is also irrelevant, as shown by the next example, which possesses the following characteristics:

- (1) Cramér's conditions and the condition of identifiability are satisfied.
- (2) The likelihood equation has roots.
- (3) With probability tending to unity, the maximum likelihood estimate exists, is unique, and is a root of the likelihood equation.
- (4) The M.L. estimate is not consistent.
- (5) There exist consistent estimates.

Let Θ be $\bigcup_k A_k$ as in the first example, and let $\{\alpha_k\}$ be an ordering of the rationals of the interval $(0, 1)$. Let $\rho(\theta) = \alpha_k$ if $\theta \in A_k$ and let (X_i, Y_i, Z_i) be multinomially distributed with probabilities $p_1 = \rho(\theta) \cos^2 2\pi\theta$, $p_2 = \rho(\theta) \sin^2 2\pi\theta$, and $p_3 = 1 - \rho(\theta)$. For n independent observations, the likelihood function is

$$\log p_n = n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + f(n_1, n_2, n_3),$$

where $n_1 = \sum_{i=1}^n X_i$, $n_2 = \sum_{i=1}^n Y_i$, $n_3 = \sum_{i=1}^n Z_i$, and f is a function which does not depend on θ . Again, the likelihood equation has solutions of the form $2\pi\theta = \tan^{-1} \sqrt{n_2/n_1}$. Since the density is maximized by taking $p_i = n_i/n$, if this is possible, only one of these solutions is the M.L. estimate. With probability tending to unity, the M.L. estimate $\hat{\theta}_n$ is such that $1 - \rho(\hat{\theta}_n) = n_3/n$. For $\hat{\theta}_n$ to be consistent, it must eventually stay in a fixed interval A_k so that $n_3/n = 1 - \alpha_k$, but the probability of this equality tends to zero as n tends to infinity.

3. Existence of consistent estimates. In the discussion of the first and second examples given above, it is stated that there exist consistent estimates. This follows from the lemma stated in the present section.

Consider a situation where the following assumptions are satisfied:

- (1) Observations are made on a sequence of independent identically distributed variables $\{X_n\}$, $n = 1, 2, \dots$, taking their values in a Euclidean space \mathfrak{X} .
- (2) The parameter space Θ is a measurable subset of a Euclidean space.
- (3) To each $\theta \in \Theta$ there corresponds a measure P_θ on \mathfrak{X} , and the distribution of the sequence $\{X_n\}$ is the product measure corresponding to a P_θ of the family $\{P_\theta, \theta \in \Theta\}$.
- (4) $P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$.
- (5) Θ is a locally compact subset of a Euclidean space and the map $\theta \rightarrow P_\theta$ is continuous in the sense that if $\theta_n \rightarrow \theta_0$, then $P_{\theta_n} \rightarrow P_{\theta_0}$ for Paul Lévy's distance.

One can easily obtain the following proposition:

LEMMA 1. *Let assumptions (1) to (5) be satisfied. Then there exists a sequence $\{T_n\}$ of estimates such that for every positive ϵ and every compact set $K \subset \Theta$ the quantity*

$$\sup_{\theta \in K} P[|T_n - \theta| > \epsilon | \theta]$$

tends to zero as n tends to infinity.

The proof of this lemma has been given elsewhere, see [5]. It entails that Cramér's conditions in [3], when supplemented by (4), above, imply the existence

of consistent estimates. Even in the simple case described by assumptions (1) to (4), the problem of finding necessary and sufficient conditions for the existence of consistent estimates has not been solved. Partial results have been obtained by C. Stein [6] and Doob [7].

4. Independent, not identically distributed, variables. In the case of independent identically distributed variables, the lemma given in Section 3 ensures the existence of consistent estimates in a wide variety of circumstances. If the variables are not identically distributed, much more freedom is available, as indicated by the next example which possesses the following characteristics:

(1) Wald's conditions [4] and the condition of identifiability are satisfied.

(2) There does not exist any consistent estimate.

Let Θ be the open interval $(0, 3\pi)$. For each θ , let X_{2i} be normal with mean $\cos a_i\theta$ and variance 1 and let X_{2i+1} be normal with mean $\sin \theta$ and variance 1. It can be verified that if a_i tends to unity, Wald's conditions for the existence of consistent roots are satisfied. However, a necessary condition for the existence of consistent estimates given in [8] is not always satisfied. In the present case, this condition would require that for any two values θ_1, θ_2 , the quantity

$$n(\sin \theta_1 - \sin \theta_2) + \sum_{i=1}^n [\cos a_i\theta_1 - \cos a_i\theta_2]^2$$

increases to infinity. If θ_2 is taken equal to $\theta_1 + 2\pi$ and the a_i 's tend to unity sufficiently fast, this condition is not satisfied, although the condition of identifiability can readily be satisfied.

It should be noted that the above is not contradictory to Wald's assertion that there is a sequence of roots which converge to the true parameter value. However there can be, as in this example, more than one limit point to the set of all roots. There is no consistent estimate because it cannot be determined from the sample values alone which convergent subsequences of the roots are the appropriate ones.

REFERENCES

- [1] A. WALD, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 595-601.
- [2] V. S. HUZURBAZAR, "The likelihood equation, consistency, and the maximum of the likelihood function," *Ann. Eugenics*, Vol. 14 (1948), pp. 185-200.
- [3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
- [4] A. WALD, "Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 40-46.
- [5] L. LECAM, "On the asymptotic theory of estimation and testing hypotheses," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, to be published.
- [6] C. STEIN, unpublished.
- [7] J. DOOB, "Applications of the theory of martingales," *Colloques Internationaux du Centre National de la Recherche Scientifique*, 1949, pp. 23-27.
- [8] C. KRAFT, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. of California Publ. Stat.*, Vol. 2, No. 6 (1955), pp. 125-

ON THE UNIQUENESS OF WALD SEQUENTIAL TESTS¹

BY LIONEL WEISS

University of Oregon

1. Summary. Under certain mild restrictions on the distributions involved, it is shown that the probabilities of the two types of error uniquely determine the two bounds characterizing the Wald sequential probability ratio test.

2. Introduction. X_1, X_2, \dots is an infinite sequence of independent and identically distributed chance variables. The density of X_i is $f_i(x)$ under H_i , where $i = 1, 2$. We assume that under either H_1 or H_2 the chance variable $f_2(X_1)/f_1(X_1)$ has a distribution which assigns a positive probability to any nondegenerate interval in the interval $[0, \infty]$, and zero probability to any point in that interval.

B, A shall denote the stopping bounds characterizing the usual Wald sequential probability ratio test. As usual, $B < A$. $Q_i[R; T]$ shall denote the probability under H_i that the value of the final probability ratio is in the region R , when the sequential stopping rule is to stop the first time the value of the probability ratio is in T , and not before; $u(z)$ shall denote the set of numbers less than or equal to z ; $v(z)$ shall denote the set of numbers greater than or equal to z . The union of any two sets R and T shall be denoted by $R + T$. We note the following easily proved inequality for future reference: if b, a are any two finite positive numbers with $b < a$, then

$$Q_2[u(b); u(b) + v(a)] < b \cdot Q_1[u(b); u(b) + v(a)].$$

In what follows, $\theta_1, \theta_2, \dots$ shall be numbers between zero and one.

For any given B, A , we denote by $\alpha(B, A)$ the probability of accepting H_2 when H_1 is true when using the Wald test with bounds B, A ; while $\beta(B, A)$ denotes the probability of accepting H_1 when H_2 is true and the Wald test with bounds B, A is used.

3. Proof of uniqueness. Let α, β be two given numbers between zero and one, such that the equalities $\alpha(B, A) = \alpha$ and $\beta(B, A) = \beta$ imply the strict inequalities $0 < B < A < \infty$. Then we have:

THEOREM. *There is at most one solution to the equations $\alpha(B, A) = \alpha, \beta(B, A) = \beta$, the unknowns being B, A .*

PROOF. We assume that there is at least one solution to these equations. Let B be any number for which it is possible to find an A greater than B with

$$\alpha(B, A) = \alpha.$$

Received August 29, 1955.

¹ Research under contract with the Office of Naval Research.

Fixing B , we shall show that the equation $\alpha(B, A) = \alpha$ is satisfied for exactly one value of A . This is so because for a fixed B , $\alpha(B, A)$ is a strictly decreasing function of A under the assumptions made above. We denote the value of A satisfying $\alpha(B, A) = \alpha$ by $A(B)$. It is easily seen that $A(B)$ is a strictly decreasing and continuous function of B , and the set of all B for which $A(B)$ exists is an interval.

From now on, we shall denote $\beta(B, A(B))$ by $\beta(B)$. Our next step is to show that $\beta(B)$ is a strictly increasing function of B , and this will complete the proof of the theorem. For a given B , we can find a positive ΔB so small that $B + \Delta B < A(B + \Delta B)$. We denote $A(B) - A(B + \Delta B)$ by ΔA . We denote by R the set of numbers no greater than B , by S the set of numbers between B and $B + \Delta B$, by T the set of numbers between $A(B) - \Delta A$ and $A(B)$, by U the set of numbers greater than $A(B)$, and finally we denote the set $R + S + T + U$ by V . We have the following relationships, where z is the variable of integration:

$$\begin{aligned} \beta(B) = & Q_2[R; V] + \int_B^{B+\Delta B} Q_2 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_2[u(z); V] \\ (3.1) \quad & + \int_{A(B)-\Delta A}^{A(B)} Q_2 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_2[u(z); V]; \end{aligned}$$

and

$$(3.2) \quad \beta(B + \Delta B) = Q_2[R; V] + Q_2[S; V].$$

Then we get

$$\begin{aligned} \beta(B + \Delta B) - \beta(B) = & Q_2[S; V] \\ (3.3) \quad & - \int_B^{B+\Delta B} Q_2 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_2[u(z); V] \\ & - \int_{A(B)-\Delta A}^{A(B)} Q_2 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_2[u(z); V]. \end{aligned}$$

Also, we have that the expression we get by replacing the subscripts 2 in the right-hand side of (3.1) by the subscripts 1 is equal to $1 - \alpha$, as is the expression on the right-hand side of (3.2) when the same change of subscripts is made. Then we get by subtraction

$$\begin{aligned}
 (3.4) \quad Q_1[S; V] &= \int_B^{B+\Delta B} Q_1 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_1[u(z); V] \\
 &\quad + \int_{A(B)-\Delta A}^{A(B)} Q_1 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_1[u(z); V].
 \end{aligned}$$

Using some obvious continuity properties of the function Q_1 , we get

$$(3.5) \quad Q_1[S; V] = (B + \theta_1 \Delta B) \cdot Q_1[S; V],$$

and combining (3.3), (3.4), and (3.5), we get

$$\begin{aligned}
 (3.6) \quad &\beta(B + \Delta B) - \beta(B) \\
 &= (B + \theta_1 \Delta B) \int_B^{B+\Delta B} Q_1 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_1[u(z); V] \\
 &\quad + (B + \theta_1 \Delta B) \int_{A(B)-\Delta A}^{A(B)} Q_1 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_1[u(z); V] \\
 &\quad - \int_B^{B+\Delta B} Q_2 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_2[u(z); V] \\
 &\quad - \int_{A(B)-\Delta A}^{A(B)} Q_2 \left[u \left(\frac{B}{z} \right); u \left(\frac{B}{z} \right) + v \left(\frac{A}{z} \right) \right] dQ_2[u(z); V].
 \end{aligned}$$

Again using continuity properties of Q_1 , we can write (3.6) as follows:

$$\begin{aligned}
 (3.7) \quad &\beta(B + \Delta B) - \beta(B) \\
 &= (B + \theta_1 \Delta B) \cdot Q_1 \left[u \left(\frac{B}{B + \theta_2 \Delta B} \right); u \left(\frac{B}{B + \theta_2 \Delta B} \right) \right. \\
 &\quad \left. + v \left(\frac{A}{B + \theta_2 \Delta B} \right) \right] \cdot Q_1[S; V] \\
 &\quad + (B + \theta_1 \Delta B) \cdot Q_1 \left[u \left(\frac{B}{A(B) - \theta_3 \Delta A} \right); u \left(\frac{B}{A(B) - \theta_3 \Delta A} \right) \right. \\
 &\quad \left. + v \left(\frac{A}{A(B) - \theta_3 \Delta A} \right) \right] \cdot Q_1[T; V] \\
 &\quad - Q_2 \left[u \left(\frac{B}{B + \theta_4 \Delta B} \right); u \left(\frac{B}{B + \theta_4 \Delta B} \right) + v \left(\frac{A}{B + \theta_4 \Delta B} \right) \right] \cdot Q_2[S; V] \\
 &\quad - Q_2 \left[u \left(\frac{B}{A(B) - \theta_5 \Delta A} \right); u \left(\frac{B}{A(B) - \theta_5 \Delta A} \right) \right. \\
 &\quad \left. + v \left(\frac{A}{A(B) - \theta_5 \Delta A} \right) \right] \cdot Q_2[T; V].
 \end{aligned}$$

But $Q_2[S; V] = (B + \theta_1 \Delta B) \cdot Q_1[S; V]$, while $Q_2[T; V] = (A(B) - \theta_5 \Delta A) \cdot Q_1[T; V]$, and using these relationships in (3.7) we get:

$$\begin{aligned}
 & \beta(B + \Delta B) - \beta(B) \\
 &= (B + \theta_1 \Delta B) \cdot Q_1[S; V] \cdot \left\{ Q_1 \left[u \left(\frac{B}{B + \theta_2 \Delta B} \right); u \left(\frac{B}{B + \theta_2 \Delta B} \right) \right. \right. \\
 & \quad \left. \left. + v \left(\frac{A}{B + \theta_2 \Delta B} \right) \right] \right. \\
 & \quad \left. - Q_2 \left[u \left(\frac{B}{B + \theta_4 \Delta B} \right); u \left(\frac{B}{B + \theta_4 \Delta B} \right) + v \left(\frac{A}{B + \theta_4 \Delta B} \right) \right] \right\} \\
 (3.8) \quad & + Q_1[T; V] \cdot \left\{ (B + \theta_1 \Delta B) \cdot Q_1 \left[u \left(\frac{B}{A(B) - \theta_2 \Delta A} \right); \right. \right. \\
 & \quad \left. \left. u \left(\frac{B}{A(B) - \theta_2 \Delta A} \right) + v \left(\frac{A}{A(B) - \theta_2 \Delta A} \right) \right] \right. \\
 & \quad \left. - (A(B) - \theta_4 \Delta A) \cdot Q_2 \left[u \left(\frac{B}{A(B) - \theta_4 \Delta A} \right); \right. \right. \\
 & \quad \left. \left. u \left(\frac{B}{A(B) - \theta_4 \Delta A} \right) + v \left(\frac{A}{A(B) - \theta_4 \Delta A} \right) \right] \right\}.
 \end{aligned}$$

Recalling that

$$\begin{aligned}
 Q_2 \left[u(1); u(1) + v \left(\frac{A}{B} \right) \right] &< Q_1 \left[u(1); u(1) + v \left(\frac{A}{B} \right) \right], \\
 Q_2 \left[u \left(\frac{B}{A(B)} \right); u \left(\frac{B}{A(B)} \right) + v \left(\frac{A}{A(B)} \right) \right] \\
 &< \frac{B}{A(B)} \cdot Q_1 \left[u \left(\frac{B}{A(B)} \right); u \left(\frac{B}{A(B)} \right) + v \left(\frac{A}{A(B)} \right) \right],
 \end{aligned}$$

and from continuity considerations on Q_1 and Q_2 , it follows that each of the two expressions in braces in (3.8) becomes positive for small enough ΔB . This proves that $\beta(B)$ is strictly increasing in B , and completes the proof of the theorem.

4. Extensions. All the results above go through in the same way under the following somewhat less restrictive conditions: Under either H_1 or H_2 , the chance variable $f_2(X_1)/f_1(X_1)$ has a continuous distribution which assigns a positive probability to any nondegenerate subinterval of $[C, D]$, where $0 \leq C < 1 < D$; and the equalities $\alpha(B, A) = \alpha$ and $\beta(B, A) = \beta$ imply the strict inequalities $C < B < A < D$.

A NOTE ON BHATTACHARYYA BOUNDS FOR THE NEGATIVE BINOMIAL DISTRIBUTION

By V. N. MURTY

University of North Carolina

In the lecture notes of Professor Lehmann on the theory of estimation [1], the first two Bhattacharyya lower bounds for the variance of an unbiased estimate of p for the negative binomial have been calculated. It is of some interest to know how the successive bounds turn out, and whether they tend to pq , which we know to be attainable. The object of the present note is to give an explicit expression for the k -th lower bound and show that it tends to pq .

If X has a negative binomial distribution, then we know that

$$(1) \quad P(X = x) = qp^x \quad x = 0, 1, 2, \dots,$$

where $q = 1 - p$. Let

$$(2) \quad S_n = \frac{1}{P(x)} \cdot \frac{\partial^n P(x)}{\partial p^n}.$$

Then it is easily verified that

$$(3) \quad S_n = \frac{(-1)^n X^{(n)}}{q^n} + \frac{(-1)^{n-1} n X^{(n-1)}}{pq^{n-1}},$$

where

$$X^{(m)} = x(x-1) \cdots (x-m+1).$$

Therefore,

$$(4) \quad S_m S_n = \frac{(-1)^{m+n}}{q^{m+n}} \left[X^{(m)} X^{(n)} - \left(\frac{q}{p}\right) m X^{(m-1)} X^{(n)} - \left(\frac{q}{p}\right) n X^{(n-1)} X^{(m)} + \left(\frac{q}{p}\right)^2 mn X^{(m-1)} X^{(n-1)} \right].$$

It is well known that

$$(5) \quad E[X^{(m)}] = m! \left(\frac{q}{p}\right)^m,$$

and we have the algebraic identity

$$(6) \quad X^{(r)} X^{(n)} = \sum_{r=0}^m \binom{m}{r} n^{(r)} X^{(n+m-r)},$$

where $n^{(r)} = n(n-1) \cdots (n-r+1)$. Therefore,

$$(7) \quad E[X^{(m)} X^{(n)}] = \sum_{r=0}^m \binom{m}{r} n^{(r)} (n+m-r)! \left(\frac{q}{p}\right)^{n+m-r}.$$

Received February 21, 1955.

Present address, Statistician, Central Tobacco Research Institute, Rajahmundry, Andhra State, India.

Using (7) in (4), after some simplification we have

$$(8) \quad E(S_m S_n) = \frac{(-1)^{m+n}}{q^m p^{n+m}} n! \sum_{r=0}^m \frac{(n+m-r-2)!}{(n-r)!} (mn-r) \binom{m}{r} p^r q^{m-r}.$$

Let $\lambda_{mn} = E(S_m S_n)$. Putting $m = 1, 2, 3$, and 4 in (8), we have in particular,

$$(9) \quad \begin{aligned} \lambda_{1n} &= \frac{(-1)^{n+1} \cdot n!}{qp^{n+1}}, \\ \lambda_{2n} &= \frac{(-1)^{n+2} \cdot n!}{q^2 p^{n+2}} [2q + 2(n-1)], \\ \lambda_{3n} &= \frac{(-1)^{n+3} \cdot n!}{q^3 p^{n+3}} [6q^2 + 12q(n-1) + 3n(n-1) + 6], \\ \lambda_{4n} &= \frac{(-1)^{n+4}}{q^4 p^{n+4}} \cdot n! [24q^3 + 72q^2(n-1) + 36q(n-1)(n-2) \\ &\quad + 4n(n+1)(n-7) + 24(3n-1)]. \end{aligned}$$

The k -th lower bound is given by

$$(10) \quad L_k = \frac{\begin{vmatrix} \lambda_{22} & \lambda_{23} & \cdots & \lambda_{2k} \\ \lambda_{32} & \lambda_{33} & \cdots & \lambda_{3k} \\ \dots & \dots & \dots & \dots \\ \lambda_{k2} & \lambda_{k3} & \cdots & \lambda_{kk} \end{vmatrix}}{\begin{vmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \dots & \dots & \dots & \dots \\ \lambda_{k1} & \lambda_{k2} & \cdots & \lambda_{kk} \end{vmatrix}}.$$

To evaluate the denominator, we multiply the first row by $2/p$ and add the second row to it; the second row is multiplied by $3/p$, and the third row is added to it; and so on. A successive application of this procedure will reduce the determinant to a triangular one, the value of which is easily computed. We thus have,

$$(11) \quad \begin{vmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \dots & \dots & \dots & \dots \\ \lambda_{k1} & \dots & \dots & \lambda_{kk} \end{vmatrix} = \frac{[k! (k-1)! \cdots 1!]^2}{q^{k(k+1)/2} p^{k(k+1)}},$$

$$\begin{vmatrix} \lambda_{22} & \cdots & \lambda_{2k} \\ \lambda_{32} & \cdots & \lambda_{3k} \\ \dots & \dots & \dots \\ \lambda_{k2} & \cdots & \lambda_{kk} \end{vmatrix} = \frac{[k! (k-1)! \cdots 1!]^2}{q^{k^2+k-2/2} p^{k^2+k-2}} \{q^{k-1} + q^{k-2} + \cdots + 1\}.$$

From (11) we have

$$L_k = p^2 q (q^{k-1} + q^{k-2} + \cdots + 1).$$

Therefore,

$$\lim_{k \rightarrow \infty} L_k = p^2 q / (1 - q) = pq.$$

REFERENCES

- [1] E. L. LEHMANN, "Notes on the theory of estimation," University of California, 1950 (mimeo).

ABSTRACTS OF PAPERS

(Additional abstracts of papers presented at the Detroit meeting of the Institute, September 7-10, 1956)

1. Further Applications of Information Theory to Multivariate Analysis and Statistical Inference, (Preliminary Report), MORTON KUPPERMAN, The George Washington University, (By Title).

A generalized statistic based on the Kullback-Leibler measure of information is defined as

$$2nI^* = 2n \int f(x, \theta^*) \log \frac{f(x, \theta^*)}{f(x, \theta_0)} d\lambda(x),$$

where the vector θ^* of h components is any consistent, asymptotically normal, efficient estimator and θ_0 is specified. $2nI^*$ is used to test the hypothesis H_0 : The sample is from a specified multivariate multiparameter population (not necessarily normal). The asymptotic distribution of $2nI^*$ under H_0 is chi-square with h d.f. I^* is modified to test the hypothesis H'_0 : r (≥ 2) samples are from the same general multivariate population, parameters not specified; its asymptotic distribution under H'_0 is chi-square with $(r-1)h$ d.f. Corresponding results are obtained for divergence-statistics based on the divergence $J(1, 2)$. Large-sample distributions of I^* and J^* under alternative hypotheses are approximated by noncentral chi-square distributions.

For any multivariate multiparameter distribution admitting sufficient statistics, $-\log \lambda = \hat{I}$, where λ is the likelihood-ratio criterion and \hat{I} uses maximum-likelihood estimators.

Information theory is applied to hypothesis testing, Pearson's chi-square test of goodness of fit, and the derivation of exact sampling distributions of sufficient statistics. It is shown that the set of sufficient estimators of population parameters appearing explicitly in any Koopman-Pitman distribution (admitting sufficient statistics) are distributed jointly in a Koopman-Pitman distribution. (Received July 19, 1956.)

2. Generalization of Thompson's Distribution, ANDRÉ G. LAURENT, Michigan State University.

Generalization of Thompson's Distribution. 1.1.) Let $X = (X_1, \dots, X_N)$ be $N(m, \sigma)$ distributed, $\bar{X} = \Sigma X_i/N$, $s^2 = \Sigma (X_i - \bar{X})^2/N$ and $t = (X_i - \bar{X})/s$. It is well known (W. R. Thompson, 1935) that $t^2/(N-1)$ is Incomplete Beta distributed and that this distribution is also the conditional distribution of any X_i , given \bar{X} and s . Three generalizations of that result are presented. 1.2.) If $\xi = (\xi_1, \dots, \xi_k)'$ is a subsample from X , the p.d.f. of $t = (\xi - \bar{X})/s$ is $[1 - t(\xi_1/N + 1/N(N-k))]t^{(N-k-1)/2} \Gamma[(N-1)/2] / \pi^{k/2} \Gamma[(N-k-1)/2] N^{(k-1)/2} (N-k)^{1/2}$. This provides also the conditional distribution of ξ , given \bar{X} and s . 2.1.) If a vector $X = (X', \dots, X'')$ is $N(m, \Sigma)$ distributed and if $\xi = (\xi', \dots, \xi'')$ is any observation from a sample $(X) = (X_1, \dots, X_N)$ 'with mean m' and covariance matrix S , the cond. p.d.f. of ξ , given m' and S , is $[1 - (\xi - m')S^{-1}(\xi - m')]'^{(N-p-1)/2} |S|^{-1/2} [(N-1)\pi]^{-1/2} \Gamma[(N-1)/2] / \Gamma[(N-1-p)/2]$. 2.2.) The latter result is generalized to the case where a subsample $(\xi) = (\xi_1, \dots, \xi_k)'$ is drawn from (X) . The conditional distribution of (ξ) , given m' and S , is a generalized multivariate Incomplete Beta distribution. These results make it possible to obtain and study the U.M.V. unbiased estimates of functions of the populations parameters, with obvious applications in the fields of S.Q.C., bombing problems, etc., and tolerance regions investigations. (Received July 23, 1956.)

3. A New Class of PBIB Designs, DALE M. MESNER, Purdue University and Michigan State University.

In partially balanced incomplete block (PBIB) designs of Latin square type with g constraints and n^2 treatments, algebraic expressions in the integers n and g give values of the parameters n_i and p_{ijk} if n and g are positive. Some negative values of n and g lead to negative parameter values which cannot occur in a design, but others give non-negative values which differ from those for any known designs and suggest the existence of a new type of design with n^2 treatments. Five families of the new designs, referred to here as the "negative Latin square" type, are constructed, based on association schemes with 16, 64, 81 (two schemes) and 100 treatments. Necessary and sufficient conditions are first given that associate classes in an association scheme may be combined to give a scheme with fewer classes. For n^2 equal to 16, 64, or 81, the new schemes are constructed by combining classes in association schemes having $n - 1$ classes of $n + 1$ treatments each, constructed from the field $GF(n^2)$ by a method similar to that of Sprott [*Can. J. Math.*, Vol. 7 (1955), 369-381]. The scheme with 100 treatments is found by an enumeration method. (Received July 23, 1956.)

4. Some Results for Inverting Patterned Matrices, A. E. SARHAN, Egyptian Medical Research Laboratories, Cairo, Egypt, and B. G. GREENBERG, University of North Carolina, (By Title).

Employing the results given in a paper by Ukita "On the Characterization of Diagonal Matrix of 2-Type and its Application to Order Statistics," generalizations can be made for various cases given in the paper by Roy and Sarhan "On Inverting a Class of Patterned Matrices," *Biometrika*, June 1956. In addition, generalizations for getting the inverse of complex matrices which can be partitioned into submatrices of the form $[D\alpha + \lambda J]$ are obtained. Matrices of this nature are frequently encountered in varied statistical applications, involving least squares, experimental design and order statistics. (Received July 23, 1956.)

5. On the Solution of the Functional Equation of Farrell's Market, A. CHARNES and O. P. AGGARWAL, Purdue University.

A. Charnes and M. Farrell in June 1953 derived a la R. Bellman a functional equation describing the optimal pricing strategy of a seller in a recurring market of Markovian reaction characteristics of "kinky oligopoly" type suggested by Farrell (*Econometrica* Vol. 22, No. 3, July 1954) in his approximate im kleinen analysis of an econometric controversy on optimality of holding price constant. They established (unpublished) existence and uniqueness of continuous solutions of at most exponential growth for this functional equation, a type still not subsumed in any yet treated (cf. S. Karlin *Naval Research Logistics Quarterly* Vol. 2, No. 4, Dec. 1955).

This paper develops explicitly the exact solution for all relevant parameter ranges. Hence (1) a new method of approximately solving "extremal" functional equations by employing piece-wise quadratic "forcing" terms is suggested and partially implemented, (2) an im grossen resolution of the econometric controversy is at hand. (Received August 29, 1956.)

6. Randomization and Experimentation, W. J. YODEN, National Bureau of Standards.

Randomization, often specified as an indispensable requirement in experimental design, is required only when the order or position of the experimental unit influences the per-

formance of the unit. Randomization, when required, may give an arrangement that is obviously undesirable and one that may doom the particular experimental program. A system of constrained randomization is proposed that eliminates the undesirable arrangements without sacrificing the customary gains achieved by randomization. (Received September 14, 1956.)

STATISTICAL RESEARCH MONOGRAPHS

The Institute of Mathematical Statistics and the University of Chicago announce the establishment and cosponsorship of a series of publications to be called *Statistical Research Monographs*.

The primary purpose of this series is to provide a medium of publication for material of interest to statisticians that is not ordinarily provided for by existing media. It will help fill the gap between journal articles and textbooks or treatises. Among the kinds of publications envisaged are

New research results too lengthy for the usual journal article. In particular, authors will have ample scope for detailed exposition of their findings.

Research results of interest in both theoretical and applied statistics. At present authors of such material frequently find it necessary to publish part of their results in a theoretical journal and part in an applied journal.

Expository monographs in particular areas of statistics.

Discussions of statistical problems and techniques in particular areas of application.

Every attempt will be made to maintain the highest standards of scholarship.

The members of the Editorial Board are David Blackwell, William G. Cochran, Henry E. Daniels, Wassily Hoeffding, Jack C. Kiefer, and William H. Kruskal (chairman). The Editor of the *Annals of Mathematical Statistics* is an ex-officio member of the Editorial Board. Members of the Editorial Board are selected by the President of the Institute and by the Committee on Statistics of the University of Chicago, subject to the approval of the Institute's Council and of the University's Board of Publications. The Editorial Board will act as a group in its decisions on manuscripts.

Institute members will receive a one-third discount on prepublication orders for monographs of the Series when such orders are placed through the Treasurer of the Institute. A smaller discount will apply to postpublication orders.

Authors are invited to send manuscripts and correspondence concerning the Series to William H. Kruskal, 127 Eckhart Hall, University of Chicago, Chicago 37, Illinois, or to any of the other members of the Editorial Board.

NEWS AND NOTICES

JOHN WISHART

Dr. John Wishart, director of the Statistical Laboratory, University of Cambridge, died in a swimming accident at the Port of Acapulco, Mexico, on July 14, 1956. As a tribute to Dr. Wishart, the Institute of Mathematical Statistics, the American Statistical Association, and the Biometric Society sponsored a memorial session on September 9, 1956, at the Detroit meetings. Professor Harold Hotelling delivered an address, "Contributions of John Wishart to Statistics."

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Sidney Addelman, who received a Master of Arts degree in statistics from the University of Delaware in June, has accepted a research assistantship from Iowa State College where he will continue his studies toward a Ph.D. degree in statistics.

F. J. Anscombe has been appointed Associate Professor in the Department of Mathematics, Princeton University.

Professor George Allen Baker of the University of California, Davis, served as Faculty Research Lecturer for 1955-56. He delivered the fourteenth annual Faculty Research Lecture, on the topic "Search for Structure."

Charles A. Bicking has accepted the position of Manager, Quality Control Branch, Research and Development Division, The Carborundum Company, Niagara Falls, New York. His responsibilities will include the establishment of process quality control in the manufacturing divisions of the company, assessment of product quality, design of experiments in research and development, and operations research.

Richard S. Bingham, Jr., has joined the newly organized Quality Control Branch, Research and Development Division, The Carborundum Company, as Senior Engineer.

Dr. Archie Blake, formerly an Advisory Engineer with Westinghouse Electric Corporation, Baltimore, has accepted an appointment as Systems Staff Mathematician with Bendix Aviation Corporation in Detroit.

Dr. Ernest E. Blanche is now president and senior research scientist of Ernest E. Blanche and Associates. The firm has offices at Rockville and Silver Spring, Maryland.

Robert C. Burton, mathematician in the Statistical Engineering Laboratory of the National Bureau of Standards, has been awarded \$300 in recognition of contributions to the theory of experimental design. The award was made possible by the "Special Act or Service" category of the Incentive Awards Act. Mr. Burton is now on educational leave from NBS, doing graduate work at the University of North Carolina.

Richard G. Cornell, who received his Ph.D. degree in statistics from Virginia Polytechnic Institute in June, is now a statistician in the Commissioned Corps of the Public Health Service. He is stationed at the Communicable Disease Center, Atlanta, Georgia.

Assistant Professor Louis J. Cote has left Purdue University to take a position at Syracuse University.

Donald P. Gaver, Jr., received the Ph.D. degree in mathematics from Princeton University in June, 1956. He is now employed by the Westinghouse Research Laboratory, Pittsburgh 35, Pennsylvania.

Dr. S. G. Ghurye has accepted an assistant professorship with the Committee on Statistics, University of Chicago, commencing in October, 1956.

William Gombert, director of the Management Engineering Department of the International Ladies' Garment Workers' Union and adjunct professor of industrial engineering at Columbia University, has been appointed professor of industrial engineering at Washington University.

Shanti S. Gupta has received his Ph.D. degree in statistics from the University of North Carolina and is now working with the Statistical Group of the Bell Telephone Laboratories, Allentown, Pennsylvania.

Dr. G. C. Helmstadter, formerly an Associate in Research of the Educational Testing Service, Princeton, New Jersey, has joined the faculty of Colorado A and M College as Psychometrist in the Office of Student Affairs and Assistant Professor in the Department of Psychology and Education.

Robert G. Hoffmann has been appointed as Statistician for the newly organized Commission on Professional and Hospital Activities, Inc.

W. H. Horton is now Manager of the Experimental Design and Statistical Analysis Section of the Materials Engineering Department, Westinghouse Electric Corporation.

Dr. Stanley Isaacson has resigned his position as a Senior Statistician with the Semiconductor Department of Westinghouse Electric Corporation in order to accept a position as Sales Manager with Gendler Stone Products Company, Des Moines, Iowa. He has also been appointed Lecturer in Statistics in the Community College of Drake University.

Eugene Lukacs has resigned from the Office of Naval Research to accept an appointment as professor of mathematics at the Catholic University of America.

John H. MacKay (Ph.D., University of North Carolina, 1956) has accepted a position as associate professor in the School of Industrial Engineering, Georgia Institute of Technology, Atlanta, where he will also do statistical work in the Engineering Experiment Station.

W. G. Madow of the University of Illinois will be at the Center for Advanced Study in the Behavioral Sciences, Stanford, California, during 1956-57.

Margaret Martin has returned to her position as Associate Professor of Biostatistics in the Department of Preventive Medicine and Public Health at Vanderbilt University after spending the academic year studying with the Committee on Statistics at the University of Chicago.

Dr. Irwin Miller, who received his Ph.D. degree in statistics at Virginia Polytechnic Institute, is now with the Applied Research Laboratory, United States Steel Corporation, in Monroeville, Pennsylvania.

Joseph M. Moser is now teaching at St. Louis University while working on a Ph.D. degree.

Dr. Mervin E. Muller of the Department of Mathematics, Cornell University, has accepted a position with the Scientific Computing Center of the International Business Machines Corporation, New York, New York.

Lt. L. M. Noel, USN, has been transferred from Princeton University to duty as executive officer of the U.S.S. Adroit operating out of Charleston, South Carolina.

Jean Roberts, former Director of Public Health Records and Statistics of the Minneapolis Health Department, is now Assistant Chief, Division of Research and Special Studies, Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

Daniel E. Sands has resigned from the position of statistician with the Squibb Institute for Medical Research to accept the position of statistician with the Applied Research Laboratory, United States Steel Corporation, Monroeville, Pennsylvania.

Professor Morris Skibinsky, of Purdue University, will spend the 1956-57 academic year at Michigan State University as Visiting Assistant Professor.

Arthur Stein, formerly with the Ordnance Communication Command at Joliet, Illinois, is now a Principal Research Engineer with the Cornell Aeronautical Laboratory in Buffalo, New York.

Zenon Szatrowski, Chairman of the Statistics Department, School of Business Administration, University of Buffalo, is on leave for the period 1955-1957. His present position is that of Staff Consultant at the Scientific Computing Center, International Business Machines Corporation. He is working on the application of electronic computers to statistical problems.

Robert J. Taylor has accepted a position as Mathematical Statistician in the Biometry Section, National Cancer Institute, NIH, Bethesda, Maryland.

Dr. John E. Walsh is now with the Military Operations Research Division, Lockheed Aircraft Corporation, Burbank, California.

Oscar Wesler, formerly Acting Assistant Professor of Statistics at Stanford University after receiving his Ph.D. degree in mathematical statistics, has been appointed Assistant Professor of Mathematics at the University of Michigan.

John W. Wilkinson received his Ph.D. degree in statistics in June, 1956 from the University of North Carolina and has accepted a position as Assistant Professor of Mathematics at Queen's University, Kingston, Ontario, Canada.

David M. G. Wishart has been appointed Lecturer in Mathematical Statistics at the University of Aberdeen, Scotland, and will take up his duties on July 1, 1956.

New Members

The following persons have been elected to membership in the Institute

May 16, 1956 to August 15, 1956

- Arnaiz Vellando, Gonzalo**, Ph.D. (Universidad de Madrid), Profesor adjunto de la Facultad de Ciencias Económicas, Universidad de Madrid, *Amnistia, 12.—Madrid, Spain.*
- Bailey, J. H.**, B.S. (University of Rhode Island), Graduate Assistant, Mathematics Department, University of Utah, *1227 East Third South, Salt Lake City, Utah.*
- Barlow, Richard Eugene**, M.A. (University of Oregon), Graduate Assistant, Department of Mathematics, University of Washington, *Seattle 5, Washington.*
- Béjar, Juan**, D.C.M. (Universidad de Madrid), Profesor de Métodos Estadísticos, Escuela de Estadística de la Universidad de Madrid, *San Bernardo No. 49, Madrid, Spain.*
- Beneš, Václav Edvard**, Ph.D. (Princeton University), Member of Technical Staff, Bell Telephone Laboratories, *Murray Hill, New Jersey.*
- Benvenuto, Andrew A.**, M.A. (Syracuse University), Graduate Teaching Assistant, Mathematics Department, University of Illinois, *115 Pleasant Street, Hartford, Connecticut.*
- Bloemen, A. R.**, M.E. (Institute of Technology, Delft, Holland), Research Fellow, Mathematical Centre, Amsterdam, Holland, *Boterdiepstraat 54^I, Amsterdam Z-2, Holland.*
- Blumenthal, Saul**, B.A. (Central High School of Philadelphia), Student, Cornell University, Sibley School of Mechanical Engineering, *313 S. 22nd St., Philadelphia 3, Pennsylvania.*
- Brennan, D. G.**, B.S. (Massachusetts Institute of Technology), Staff Member, Lincoln Laboratory and Graduate Student, M.I.T., *300 Westgate West, Cambridge 39, Massachusetts.*
- Brown, D. M.**, B.A. (University of Toronto), Graduate Student, University of Toronto, *73 St. George Street, Toronto, Ontario, Canada.*
- Calhoun, Carolyn M.**, B.A. (University of Alabama), Graduate Student with Assistantship, Psychology Department, University of Alabama, *Box 5331, University, Alabama.*
- Campbell, Loudon Lee**, B.S. (University of Pittsburgh), District Coordinator, Parke, Davis and Company, Department of Clinical Investigation, *Joseph Campau at the River, Detroit, Michigan.*
- Carlson, C. Henry**, M.A. (University of Illinois), Research Analyst, Douglas Aircraft Company, Santa Monica, California, *866 Haverford, Apt. 10, Pacific Palisades, California.*
- Carpenter, J. A.**, M.A. (University of North Carolina), Research Engineer, Melpar Inc., Boston, Massachusetts, *10 Forest St., Apt. 16, Cambridge 40, Massachusetts.*
- Carr, Charles, R. M.I.A.** (Columbia University), Graduate Student and Teaching Assistant, Departments of Statistics and Economics, Stanford University, *2084 Harvard Street, Palo Alto, California.*
- Chapman, Herman Hollis**, Ph.D. (Columbia University), Professor of Business Statistics, University of Alabama, *University, Alabama.*
- Cohen, Leonard**, B.S.S. (College of the City of New York), Teaching Lecturer, College of the City of New York, *1055 Wheeler Avenue, Bronx 72, New York.*
- Collier, Raymond Oliver, Jr.**, Ph.D. (University of Minnesota), Assistant Professor of Education, University of Minnesota, *1884 N. Oxford St., St. Paul 13, Minn.*
- Constantine, Alan Graham**, B.S. (University of Western Australia), Research Officer, Division of Mathematical Statistics, C.S.I.R.O., *36 Hardy St., Canterbury, N.S.W., Australia.*
- Crystal, Eugene**, A.B. (William Jewell College), Mathematician in charge of computing laboratory, Textile Research Institute, Princeton, New Jersey, *25 Witherspoon Street, Princeton, New Jersey.*
- Cuttle, Yvonne, G. M. G. (Mrs. P. M.)**, M.A. (University of British Columbia), Graduate Student, University of Oregon, *Eugene, Oregon.*

- Dalton, Jonas, M., A.B.** (George Washington University), Graduate Student, Virginia Polytechnic Institute, Blacksburg, Virginia, *204 Roanoke St., Blacksburg, Virginia.*
- Davis, Willis L., B.S.** (Howard University), Acting Group Leader, Statistical Services, Research and Advanced Development Division, AVCO Manufacturing Corporation, Stratford, Connecticut, *1501 Seaview Avenue, Bridgeport, Connecticut.*
- Dick, Ronald S., B.S.** (Queen's College), Assistant in Mathematical Statistics, Columbia University, and Mathematical Statistician, U.S. Census Bureau, Washington, D. C., *84-66 250th St., Bellerose 26, New York.*
- Dresner, A. Joseph, M.A.** (New York University), Statistician, Board of Education, City of New York *161-30 Jewel Ave., Flushing 65, Long Island, New York.*
- Dunn, Olive Jean (Mrs. R. L.), M.A.** (University of California at Los Angeles), Teaching Assistant, U.C.L.A. School of Business Administration, *404 9th St., Manhattan Beach, California.*
- Eeden, Constance van, dra** (University of Amsterdam), Research Fellow, Mathematical Centre, Statistical Department, *Edisonstraat 9^{II}, Amsterdam, Holland.*
- Elmaghraby, Salah Eldin, A., M.Sc.** (Ohio State University), Research Assistant, Cornell University, and Graduate Student, Industrial Engineering, Cornell University, *409 Eddy St., Ithaca, New York.*
- Elteren, Ph. van, M.S.** (University of Amsterdam), Sub-chief of the Statistical Department, Mathematical Centre, *Ritzema Bosstraat 38^I, Amsterdam O, Netherlands.*
- Ferrer Martin, Sebastian, L.C.E.** (Universidad de Madrid), Profesor de Técnica del Muestreo, Jefe de la Sección de Metodología, Madrid, *Sagunto, 3.—Madrid, Spain.*
- Fetters, William B., B.S.** (Indiana University), Analytical Statistician, U.S. Naval Powder Factory, Quality Control Department, Indian Head, Maryland, *122 Seneca Drive, Forest Heights, Maryland.*
- Fossum, R. R., M.S.** (University of Oregon), Research Fellow, Department of Mathematics, University of Oregon, *Electronic Defense Laboratory, P.O. Box 205, Mountain View, California.*
- Fukuda, Yoichiro, M.A.** (University of California at Los Angeles), Research Mathematician, Engineering Department, University of California at Los Angeles, Los Angeles 24, California.
- Gafarian, A. V., B.S.** (University of Michigan), Student and Teaching Assistant, Department of Mathematics, University of California at Los Angeles, *5386 Daves Ave., Culver City, California.*
- Gani, J. M., Ph.D.** (Australian National University), Nuffield Research Fellow, Statistical Laboratory, The University of Manchester, Manchester 13, England.
- Gardiner, Donald A., Ph.D.** (North Carolina State College), Member, Mathematics Panel, Oak Ridge, National Laboratory, Post Office Box P, Oak Ridge, Tennessee.
- Gentry, Charles J., B.S.** (University of Florida), Graduate Student, University of Florida, *Box 2003, University Station, University of Florida, Gainesville, Florida.*
- Geoghagen, Randolph R. M., M.A.** (Columbia University), Research Assistant, Popin Cyclotron Laboratory, Columbia University, *319 East 162nd St., Bronx 56, New York.*
- Gephart, L. S., Ph.D.** (University of Florida), Mathematical Statistician, Design of Experiment Unit, Research and Development Division, Office, Chief of Ordnance, D. A., Washington 25, D. C.
- Gershefski, George William, B.M.E.** (General Motors Institute), Student, Cornell University, *127 College Avenue, Ithaca, New York.*
- Gessford, John E., M.S.** (Stanford), Graduate Student, Statistics Department, Stanford University, *637 Alvarado, Stanford, California.*
- Gorman, T. P., M.S.** (Fordham College), Senior Programmer, New York Scientific Computing Center, International Business Machines, *590 Madison Ave., New York, New York.*

- Groves, John Ellis, Jr., B.S. (Arkansas State College), Graduate Fellow, St. Louis University, 359 N. Whittier, St. Louis, Missouri.
- Halter, Albert N., Ph.D. (Michigan State University), Graduate Student and Research Instructor, Michigan State University, 2276 Hulett Road, Okemos, Michigan.
- Haynam, George E., M.S. (Case Institute of Technology), Research Assistant, Project Doanbrook, Case Institute of Technology, 10900 Euclid Avenue, Cleveland, Ohio.
- Hemelrÿk, F., Dr.Sc. (University of Amsterdam), Professor in Statistics, Technical University, Delft, and Chief of Statistical Consultation, Mathematical Centre, Amsterdam, Weesperzijde 83^{aa}, Amsterdam (O), Netherlands.
- Hetz, Wolfgang, Ph.D. (University of Göttingen), Adviser in Statistical Problems, Deutsche Forschungsgemeinschaft, Calsostr. 20, Göttingen, Germany.
- Hogben, David, B.A. (University of Minnesota), Inspection Development Engineer, Western Electric Co., 100 Central Ave., Kearny, New Jersey.
- Howes, David R., A.B. (Amherst), Staff Statistician, Chemical Corps Engineering Command, Army Chemical Center, Maryland.
- Ikedai, Hiroshi, B.P. (University of Tokyo), Graduate Student, Division of Research in Humanities, University of Tokyo, 845 Izumichô, Suginamiku, Tokyo, Japan.
- Kahn, Paul Markham, B.S. (Stanford), Student, Stanford University, 76 Manzanita Road, Fairfax, California.
- Keys, Phillis Allen, B.S. (Wayne University), Computer, Ames Laboratory, Atomic Energy Commission, and Graduate Student, Statistics Department, Iowa State College, 510 Forest Glen, Ames, Iowa.
- Kopka, William E., M.A. (Syracuse University), Teaching Assistant, Mathematics Department, Syracuse University, Syracuse 10, New York.
- Lechner, J. A., B.S. (Carnegie Institute of Technology), Graduate Student, Princeton University, Fine Hall, Box 708, Princeton, New Jersey.
- Leiman, John M., Ph.D. (University of Washington), Chief, Statistical Methodology and Analysis Branch, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas.
- Lenthall, Jerry, B.A. (Swarthmore College), Graduate Assistant, Department of Statistics, Stanford University, Box 1431, Stanford, California.
- Lieberman, Alfred, M.A. (University of Southern California), Mathematical Statistician, Bureau of Ships, Department of the Navy, Washington 25, D. C., 6019 Strathmore Ave., Kensington, Maryland.
- Lindgren, B. W., Ph.D. (University of Minnesota), Mathematics Instructor, University of Minnesota, 32 Orlin Avenue S.E., Minneapolis 14, Minnesota.
- McCue, Edmund B., M.S. (University of Michigan), Graduate Assistant, State University of Iowa, Iowa City, Iowa, 306 East Jefferson St., Iowa City, Iowa.
- McLynn, James M., M.A. (George Washington University), Research Analyst, Army Logistics Project GWV, Washington, D. C., 7417 17th Ave., West Hyattsville, Maryland.
- Mackropoulos, C. L., M.A. (University of Illinois), Assistant, Mathematics Department, University of Illinois, % Mrs. Mary E. Bourke, 7724 South Marquette, Chicago, Illinois.
- Magness, T. A., M.A. (University of California at Los Angeles), Mathematician, Ramo-Wooldridge Corporation, 8820 Bellanca Ave., Los Angeles 45, California, 1042½ S. Serano Ave., Los Angeles 8, California.
- Malan, D. J., M.A. (Columbia University), Assistant Actuary, South African National Life Assurance Company, Sanlamhof, C.P., Union of South Africa, % SANLAM, Sanlamhof, C.P., Union of South Africa.
- Matthes, T. K., B.S. (California Institute of Technology), Graduate Student, California Institute of Technology, 1201 E. California St., Pasadena 4, California.
- Merrill, Robert B., B.S. (Purdue University), Quality Control Statistician, Whirlpool-Seegar Corporation, Estate Division, Hamilton, Ohio, 5727 Marmion Lane, Cincinnati 13, Ohio.

- Miller, D. D., B.A. (University of Florida), Graduate Assistant, Department of Economics, University of Florida, *Box 2793, University Station, Gainesville, Florida.*
- Mills, Donald F., M.E. (University of Washington), Graduate Student, University of Washington, *5015 16th Ave., N. E., Seattle 5, Washington.*
- Mohan, Chandra, M.Sc. (Agra University), Research Scholar, Columbia University, % Mr. H. D. Khandelwal, *54 Lanercost Road, London S. W. 2, England.*
- Morgenthaler, George W., Ph.D. (University of Chicago), Group Leader, Institute for Air Weapons Research, University of Chicago, *3827 Woodside Avenue, Hollywood, Illinois.*
- Moursund, Andrew F., Ph.D. (Brown University), Professor and Head of Department of Mathematics, University of Oregon, Eugene, Oregon.
- Neuwirth, Lee, M.A. (Columbia University), Lecturer, Columbia College, *Box 45, Hamilton Hall, Columbia College, Morningside Heights, New York, New York.*
- Oliveira, Augusto T., B.A. (Portuguese Agricultural Research Station), Statistician, Portuguese Agricultural Research Station, *Estação Agronômica Nacional, Saca veur (Lisbon), Portugal, 676 Pammel Court, Ames, Iowa.*
- Otsuka, Jun, M.A. (Tokyo University), Student of Doctor Course, Laboratory of Animal Anatomy, Faculty of Agriculture, Tokyo University, *No. 17, 2-Chome, Kaname-cho, Toshima-ku, Tokyo-to, Japan.*
- The Parke Mathematical Laboratories, Inc. Applied Mathematicians, has moved from Concord to larger quarters in the country, *Bedford Road, Carlisle, Massachusetts.*
- Perry, N. C., Ph.D. (University of Southern California), Associate Professor, Alabama Polytechnic Institute, *135 Cedar Crest Drive, Auburn, Alabama.*
- Rappaport, Erle, M.S. (University of Michigan), Project Statistician, Aeronautical Radio, Inc., Washington, D. C., *713 Center Hunting Towers, Alexandria, Virginia.*
- Reed, Frank C., B.S. (University of Redlands), Graduate Student, University of California at Los Angeles, Westwood, California, *6150 S. Sepulveda, Culver City, California.*
- Rodine, Robert H., B.S. (State Teacher's College, Pennsylvania), Graduate Research Assistant, Statistical Laboratory, Purdue University, Lafayette, Indiana.
- Rogers, Gerald S., M.A. (University of Washington), Half-time Instructor, State University of Iowa, Iowa City, Iowa.
- Runnenburg, J. Theo, Ph.D. (University of Amsterdam), Co-worker, Statistics Department, Mathematical Centre, Amsterdam, *Overtoom 417, Amsterdam (West), Holland.*
- Russell, Thomas S., Ph.D. (Virginia Polytechnic Institute), Assistant Professor, Virginia Polytechnic Institute, *1415 Burruss Blvd., Blacksburg, Virginia.*
- Selig, Seymour M., B.M.E. (Rensselaer Polytechnic Institute), Engineering Statistician, Staff Statistician, Chemical Corps Engineering Command, Army Chemical Center, Maryland.
- Seltzer, Frederic, B.S. (City College of New York), Actuarial Student, Metropolitan Life Insurance Company, New York City, *23 Kerrigan St., Long Beach, New York.*
- Smith, Robert L., M.S. (Virginia Polytechnic Institute), Graduate Student, Virginia Polytechnic Institute, *University Club, Blacksburg, Virginia.*
- Smith, Thaddeus L., M.A. (Columbia University), *625 W. 135th Street, New York, New York.*
- Stanley, Julian C., Jr., Ed.D. (Harvard University), Associate Professor of Education, University of Wisconsin, *305 S. Owen Drive, Madison, Wisconsin.*
- Stinson, Fannie A., M.S. (Howard University), Mathematician, U. S. Navy Hydrographic Office, Washington 25, D. C., *3816 10th Street, N. W., Washington 11, D. C.*
- Takeuchi, Kei, B.E. (Tokyo University), Graduate Student, Tokyo University, *4-552 Maba shi, Suginami-ku, Tokyo, Japan.*
- Testerman, Jack, B.A. (Oklahoma A. and M. College), Graduate Assistant, Oklahoma A. and M. College, Stillwater, Oklahoma, *115 E. Boeing Drive, Midwest City, Oklahoma.*
- Thomas, Earl A., B.S. (Columbia University), Senior Reliability Analyst, Research and Advanced Development Division, AVCO Manufacturing Corporation, Stratford, Connecticut, *55 Ocean Avenue, Lordship, Connecticut.*

- Thomas, G. E., Jr.**, Ph.D. (Cornell University), Associate Professor and Executive Officer, Department of Mathematics, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.
- Urbanik, John G.**, A.B. (University of Rochester), Systems Engineer, Republic Aviation Corporation, Farmingdale, New York, 108 Rockville Centre Parkway, Oceanside, New York.
- Ury, Hans K.**, A.B. (University of California), Student and Research Assistant, Department of Statistics, University of California, Berkeley, California.
- Ward, Joe H., Jr.**, Ph.D. (University of Texas), Research Psychologist, Air Force Personnel and Training Research Center, Personnel Research Laboratory, Lackland Air Force Base, San Antonio, 315 Palm Drive, San Antonio, Texas.
- Wiesen, J. M.**, M.S. (Iowa State College), Supervisor, Statistical Division, Sandia Corporation, Albuquerque, 1609 Cagua Drive, N. E., Albuquerque, New Mexico.
- Yonezawa, Shingo**, B.E. (University of Tokyo), Graduate Student, Department of Applied Physics, Faculty of Engineering, University of Tokyo, Hongo, Tokyo, Japan.
- Yukihide, Okano**, B.E. (Tokyo University), Graduate Student, Tokyo University, 56 Otsuka-machi Bunkyo-ku, Tokyo, Japan.
- Zetterberg, Lars-Henning**, L.T. (Royal Institute of Technology, Stockholm), Research Engineer, Research Institute of National Defense, Stockholm, Sweden, 5536 Dorchester Ave., Chicago 37, Illinois.
- Zoellner, J. Arthur**, M.S. (Iowa State College), Experiment Design and Analysis Statistician, General Engineering Laboratory, General Electric Company, Schenectady, New York.
- Zolczynski, Stephen J.**, B.S. (University of Alabama), Tabulation Project Planner, Officer Education Research Laboratory, AFPTRC, Maxwell Air Force Base, Alabama, 2228 St. Charles Avenue, Montgomery, Alabama.
- Zorua Terol, Procopio**, D.C.M. (Universidad de Madrid), Profesor adjunto de Estadística Matemática, Universidad de Madrid, Avda. Aureliano Ibarra, 1, Alicante, Spain.

Preparation of Concise Tables for Statisticians Now Underway

A Team of statisticians and computers is working on a research project organized and directed by Dr. K. C. S. Pillai, UN Senior Statistical Advisor to the Statistical Center, University of the Philippines. The project involves the preparation of some new statistical tables which will be useful for tests of hypotheses in multivariate analysis. These tables will facilitate tests of equality between variate means of several multivariate populations. They will also be useful in tests of equality of dispersion matrices in two multivariate populations, and of the independence of two sets of variates which follow the multivariate normal law. It is hoped that these tables may be completed shortly to render them available to statisticians within the current year. These tables, along with some others used for tests dealing with univariate problems, will be issued by the Statistical Center in a report edited by Dr. Pillai and entitled *Concise Tables for Statisticians*.

Postdoctoral Study in Statistics

Awards for study in statistics by persons whose primary field is not statistics but one of the physical, biological, or social sciences to which statistics can be

applied are offered by the Committee on Statistics of the University of Chicago. The awards range from \$3,600 to \$5,000 on the basis of an eleven month residence. The closing date for application for the academic year 1957-8 is February 15, 1957. Further information may be obtained from the Committee on Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois.

Preliminary Actuarial Examinations Prize Awards

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of Part 2 of the 1956 Preliminary Actuarial Examination are as follows:

First Prize of \$200

Pratt, Richard L. Washington University

Additional Prizes of \$100 each

Brillinger, David R. University of Toronto

Earle, Clifford J., Jr. Swarthmore College

Kaplan, Stanley Cornell University

Mosher, Robert E. Kenyon College

Riehm, Carl R. University of Toronto

Rubin, Jerrold Columbia University

Schweitzer, Paul A. Holy Cross College

Soderquist, George D. Drake University

The Society of Actuaries has authorized a similar set of nine prizes for the 1957 examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations:

Part 1. Language Aptitude Examination.

(Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning.)

Part 2. General Mathematics Examination.

(Algebra, trigonometry, coordinate geometry, differential and integral calculus.)

Part 3. Special Mathematics Examination.

(Finite differences, probability and statistics.)

The 1957 Preliminary Actuarial Examinations will be prepared by the Educational Testing Service under the direction of a committee of actuaries and mathematicians and will be administered by the Society of Actuaries at centers throughout the United States and Canada on May 15, 1957. The closing date for applications is April 1, 1957.

The Society of Actuaries
208 South LaSalle Street
Chicago 4, Illinois

Quality Control and Applied Statistics Abstract Service

Interscience Publishers, Inc. announces the inauguration of **QUALITY CONTROL AND APPLIED STATISTICS ABSTRACTS**, a *monthly loose-leaf abstract service* covering the world literature on Quality Control, Operations Research and Industrial Applications of Statistical Methods of all kinds. More than 400 journals will be scanned, for articles that present new information in the field, and the abstracts will be sufficiently comprehensive to show the significant contribution of each article, so that it will usually be unnecessary to consult the original paper. It will consist of one volume of about 1,000 pages yearly, divided among 12 issues, beginning in June 1956. The subscription price is \$60.00 per volume.

Research Fellowships in Psychometrics

Princeton, N. J.: The Educational Testing Service is offering for 1957-58 its tenth series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of \$2,500 a year and are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School. Suitable undergraduate preparation may consist either of a major in psychology with supporting work in mathematics, or a major in mathematics together with some work in psychology. However, in choosing fellows, primary emphasis is given to superior scholastic attainment and demonstrated research ability rather than to specific course preparation. The closing date for completing applications is January 4, 1957. Information and application blanks will be available about October 1 and may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

Summer Job Opportunities at the National Bureau of Standards

The Junior Scientist-Engineer program, open primarily to sophomores and juniors, is designed to prepare especially well qualified students majoring in the physical sciences, mathematics and engineering, for a future professional career at the National Bureau of Standards. Approximately 160 college students participate in this summer program each year with approximately 40 universities represented in the group. The program is a work-study plan. Students who meet program requirements are carried on a "leave without pay" status during the school year. With the exception of the GS-1 level, selections are made from Civil Service registers which are established as a result of the Student Trainee examination.

The GS-1 group consists of selected high school graduates who have distinguished themselves in the physical sciences and engineering via the Westinghouse Science Talent Search or Science Honors on a national basis. The students participate in a planned program of orientation carefully supervised on-the-job training assignments, and discussions with trainee advisors who are appointed from each technical division.

Students interested in this program should watch for the Civil Service Commission's Student Trainee examination announcement generally posted on college bulletin boards sometime during fall or early winter.

NRC-NBS Research Associateships

Research associateships, supported by the National Bureau of Standards and awarded on recommendations of the National Academy of Sciences—National Research Council, are offered to provide young investigators of unusual promise and ability the opportunity for basic research in various branches of the physical and mathematical sciences. These associateships are open only to citizens of the United States and are tenable at the National Bureau of Standards in Washington, D. C. Applicants must have the Ph.D. or Sc.D. degree, or their equivalent. The term of the appointment is for one calendar year. It is expected that approximately 10 awards may be made in a total of fourteen fields, of which the following are of particular interest to mathematicians: Pure and Applied Mathematics; Applied Mathematical Statistics; Numerical Analysis; Statistical Mechanics. Awards will be made about April 1, 1957. Appointments will be for one year. The annual gross stipend will be \$7035 and will be subject to income tax. Requests for application forms and for additional information about requirements for applications should be addressed to the Fellowship Office, National Academy of Sciences—National Research Council, 2101 Constitution Avenue, N.W., Washington 25, D. C. *Applications for the academic year 1957-1958 must be received in the Fellowship Office no later than January 11, 1957.*

International Travel Announcement

The National Science Foundation will award individual grants to defray partial travel expenses for a limited number of American scientists participating in the following international congresses: 30th Session of the International Statistical Institute; Congress of the International Union of the Scientific Study of Population. These congresses are scheduled to meet in Stockholm, Sweden, August 8 to 15, 1957. Application blanks may be obtained from the National Science Foundation, Washington 25, D. C. *Completed application forms must be submitted by March 1, 1957.*

REPORT OF THE DETROIT MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The seventy-second meeting of the Institute of Mathematical Statistics, a Central Regional Meeting, was held at the Hotels Sheraton-Cadillac and Statler, Detroit, Michigan, on September 7-9, 1956. The meeting was in conjunction with meetings of the American Statistical Association, the American Sociological Society, the Econometric Society, the Biometric Society, the Federation of Financial Analysts Societies, the Society for the Study of Social Problems, and the Rural Sociological Society.

A Special Invited Paper, *Randomization and Industrial Experimentation*, was presented by Dr. W. J. Youden, National Bureau of Standards.

The following members of the Institute attended the Detroit meeting:

Om P. Aggerwal, Wm. R. Allen, R. L. Anderson, Theo. W. Anderson, Jr., Virgil L. Anderson, Wm. B. Anderson, Kenneth J. Arnold, Wm. Dowell Batten, Z. W. Birnbaum, Chester I. Bliss, Colin Ross Blyth, Helen Bozovich, Ralph Allan Bradley, Alva Esmond Brandt, Glenn W. Brier, Harold F. Bright, Irwin D. S. Bross, Byron Wm. Brown, Robert W. Burgess, Irving W. Burr, Elizabeth Dean Bushell, Joseph M. Cameron, Mavis B. Carroll, Victor Chew, Williard H. Clatworthy, Alonzo C. Cohen, Samuel E. Cohen, Wm. Stokes Connor, Louis J. Cote, Dudley J. Cowden, Gertrude M. Cox, Cecil C. Craig, Joseph F. Daly, Herbert T. David, Besse B. Day, Claude DeCourval, Daniel B. DeLury, Francis R. Del Priore, Lucile Derrick, Stuart C. Dodd, James L. Dorby, Acheson J. Duncan, Chas. W. Dunnett, Paul S. Dwyer, Marjorie Easterbrook, Churchill Eisenhart, Benjamin Epstein, Chas. F. Federspiel, Wm. Brooke Feters, Lester R. Frankel, Spencer M. Free, Jr., Fred Frishman, Donald A. Gardiner, Norman R. Garner, Seymour Geisser, Dorothy Morrow Gilford, Leo A. Goodman, Samuel W. Greenhouse, Joseph A. Greenwood, Lee Gunlogson, Paul Gunther, Keet W. Halbert, Max Halperin, Albert N. Halter, Morris H. Hansen, Boyd Harshbarger, H. Leon Harter, Herman O. Hartley, Wm. C. Healy, Jr., F. M. Hemphill, Leon H. Herbach, Irene Hess, Clifford G. Hildreth, Robt. G. Hoffman, Robert Hooke, Wm. H. Horton, Daniel G. Horvitz, Harold Hotelling, Earl E. Houseman, Hendrik Houthakker, Cyril C. Hoyt, Paul E. Irick, J. Edward Jackson, Palmer O. Johnson, Howard L. Jones, Lawrence F. Jones, Hyman B. Kaitz, Edward L. Kaplan, Marvin Kastenbaum, Leo Katz, Harriet J. Kelley, Lester S. Kellogg, Oscar Kempthorne, Robt. W. Kennard, George H. Kennedy, Allyn W. Kimball, Arnold J. King, Edward P. King, Calvin J. Kirchen, Leslie Kish, Carl F. Kossack, Robt. M. Kozelka, Clyde Y. Kramer, Wm. H. Kruskal, T. T. Kwo, Donald E. Lamphiear, James F. Lanahan, Andre G. Laurent, Alfred Lieberman, Morris M. Lightstone, Rensis Likert, Benjamin Lipstein, Geo. F. Lunger, Albert Madansky, G. L. Marcus, Eli S. Marks, Robt. H. Matthias, Paul Meier, Dale M. Mesner, Irwin Miller, Albert Mindlin, Joseph E. Morton, Jack Moshman, Hugo Muench, Mervin E. Muller, John Neter, Monroe L. Norden, Horace W. Norton, James A. Norton, Edwin G. Olds, Ingram Olkin, Paul S. Olmstead, Thos. M. Oneson, Bernard Ostle, Donald B. Owen, Wm. R. Pabst, Nancy S. Parker, John F. Pauls, Kan-Chen Peng, Eugene W. Pike, James H. Powell, John W. Pratt, Bruce P. Price, Donald L. Richter, David Rubenstein, Daniel E. Sands, F. E. Satterthwaite, Edward Sax, Marvin A. Schneidman, Norman C. Severo, Richard H. Shaw, Walt R. Simmons, Morris Skibinsky, H. Fairfield Smith, Paul N. Somerville, Frederick F. Stephan, John N. Stewart, Fred L. Strodbeck, Seiji Sugihara, Zenon Szatrowski, Daniel Teichroew, James G. C. Templeton, Benjamin J. Tepping, Milton E. Terry, Donovan J. Thompson, Wm. A. Thompson, Geo. Wm. Thomson, Leo J. Tick, Chia Kuei Tsao, John W. Tukey, Malcolm E. Turner, Elizabeth Vaughan, Joseph Waksberg, Louis Weitter, Alfred G. Whitney, John M. Wiesen, Gregory P. Williams, R. Lowell Wine, Gerald Winston, Wm. W. Wolman, Max A. Woodbury, Wm. J. Youden, Marvin Zelen, John A. Zoellner.

The program of the Detroit meeting follows:

FRIDAY, SEPTEMBER 7, 1956

10:00 a.m. Covariance Analysis, I. Cosponsored with the American Statistical Association and the Biometric Society.

Chairman: A. W. KIMBALL, Oak Ridge National Laboratory.

Papers: *Elements of Covariance*, D. B. DELURY, Ontario Research Foundation

Interpretations of Regressions in Analysis of Covariance, H. FAIRFIELD SMITH,
North Carolina State College.

Discussion: OSCAR KEMPTHORNE, Iowa State College.

1:00 p.m. Contributed Papers.

Chairman: JAMES LANAHAN, University of Detroit.

Papers: 1. *Some Results on the Distribution of the Peaks of a Gaussian Process*, IRWIN MILLER AND JOHN E. FREUND, Virginia Polytechnic Institute.

2. *Unbiased Estimation of the Normal Distribution Function. (Preliminary report)*, WILLIAM C. HEALY, JR., Ethyl Corporation Research Laboratories, Ferndale, Michigan.

3. *Unbiased Estimation of Correlation Coefficients*, INGRAM OLKIN AND JOHN W. PRATT, University of Chicago. (By title)

4. *On a Multivariate Tchebycheff Inequality. (Preliminary report)*, INGRAM OLKIN AND JOHN W. PRATT, University of Chicago. (By title)

5. *A Continuous Time Treatment of the Waiting-time in a Queueing System Having Poisson Arrivals, a General Distribution of Service-time, and a Single Service Unit. (Preliminary report)*, VACLAV EDVARD BENEŠ, Bell Telephone Laboratories, Murray Hill, New Jersey. (By title)

6. *Some Results on the Analysis of Random Signals by Means of a Cut-counting Process*, IRWIN MILLER AND JOHN E. FREUND, Virginia Polytechnic Institute. (By title)

7. *A New Class of Partially Balanced Incomplete Block Designs*, DALE M. MESNER, Purdue University and Michigan State University.

8. *Generalization of Thompson's Distribution*, ANDRE G. LAURENT, Michigan State University.

9. *Some Results for Inverting Patterned Matrices*, A. E. SARHAN AND B. G. GREENBERG, University of North Carolina. (By title)

10. *On the Solution of the Functional Equation of Farrell's Market*, A. CHARNES AND O. P. AGGARWAL, Purdue University. (By title)

2:00 p.m. Covariance Analysis, II. Cosponsored with the American Statistical Association and the Biometric Society.

Chairman: GERTRUDE M. COX, North Carolina State College.

Papers: *Covariance Analysis with Unequal Subclass Numbers*, WALTER T. FEDERER, Cornell University.

The Analysis of Covariance for Incomplete Block Designs, MARVIN ZELEN, National Bureau of Standards.

Group Comparisons and Analysis of Variance and Covariance in Cluster Sampling, H. O. HARTLEY, Iowa State College.

Discussion: JOHN W. TUKEY, Princeton University.

4:00 p.m. Applications of Electronic Computers in Statistics. Cosponsored with the American Statistical Association and the Biometric Society.

Chairman: M. A. WOODBURY, New York University.

Papers: *Experiences with SEAC*, J. M. CAMERON, National Bureau of Standards.

How to Control the Digital Computer, R. W. HAMMING, Bell Telephone Laboratories.

Sampling Experiments, D. TEICHROEW, National Cash Register Company.

Discussion: ZENON SZATROWSKI, International Business Machines Corporation, New York.

8:00 p.m. Theoretical Aspects of Sample Surveys. Cosponsored with the American Statistical Association.

Chairman: HOWARD L. JONES, Illinois Bell Telephone Company.

Papers: *Estimation of Variances and Composite Estimation Procedures*, JOSEPH WAKSBERG, Bureau of the Census.

Unsolved Problems in the Statistics of Survey Sampling, LESLIE KISH, University of Michigan.

Unbiased Ratio Estimators and Their Variances, LEO GOODMAN, University of Chicago, AND H. O. HARTLEY, Iowa State College.

Discussion: D. HORVITZ, North Carolina State College. ALLEN ROSS, State University of New York.

SATURDAY, SEPTEMBER 8, 1956

9:00 a.m. Critical Problems in New Quantitative Techniques. Cosponsored with the American Statistical Association and the American Sociological Society.

Chairman: LEO A. GOODMAN, University of Chicago.

Papers: *Stochastic Models and their Applications to Social Phenomena*, JERZY NEYMAN, University of California, AND WILLIAM KRUSKAL, University of California and University of Chicago.

Measurement and Sampling in Social Research, FREDERICK F. STEPHAN, Princeton University.

Discussion: PAUL F. LAZARSFELD, Columbia University.

11:00 a.m. Special Invited Paper.

Chairman: WILLIAM KRUSKAL, University of Chicago.

Address: *Randomization and Industrial Experimentation*, W. J. YODEN, National Bureau of Standards.

2:00 p.m. Invited Papers on Mathematical Statistics, I.

Chairman: LEO A. GOODMAN, University of Chicago.

Papers: *The Use of Sample Spacing in Tests of Fit*, LIONEL WEISS, University of Virginia and University of Oregon.

A Generalization of Internal Regression for the Fitting of Some Non-linear Models, R. WHITE AND O. KEMPTHORNE, Iowa State College.

On the Relative Position of the Mean and Order Statistics, H. T. DAVID, University of Chicago and Iowa State College.

SUNDAY, SEPTEMBER 9, 1956**10:00 a.m. Applications of Stochastic Processes. Cosponsored with the American Statistical Association and the Biometric Society.**

Chairman: OSCAR KEMPTHORNE, Iowa State College.

Papers: *The After-History of Pulmonary Tuberculosis, A Stochastic Model*, DAVID W. ALLING
Herman W. Biggs Memorial Hospital.

The Application of Stochastic Processes to the Kinetics of Enzyme Action, ANTHONY
F. BARTHOLOMAY, Harvard School of Public Health and Harvard University
Medical School.

2:00 p.m. Wishart Memorial Session. Cosponsored with the American Statistical Association and the Biometric Society.

Chairman: GERTRUDE COX, North Carolina State College.

Address: *Contributions of John Wishart to Statistics*, HAROLD HOTELLING, University of
North Carolina.

4:00 p.m. Invited Papers on Mathematical Standards, II.

Chairman: LOUIS J. COTE, Syracuse University.

Papers: *Decision Theory for Polya Type Distributions*, JOHN PRATT, University of Chicago.
Bayes Two-stage Decision Rules, MORRIS SKIBINSKY, Purdue University.

The Chairman of the Program Committee for the meeting was Stanley Isaacson
Des Moines, Iowa. The Assistant Secretary for the meeting was James Lanahan,
University of Detroit.

WILLIAM KRUSKAL
Associate Secretary

REPORT OF THE SEATTLE MEETING OF THE INSTITUTE

The seventy-first meeting of the Institute of Mathematical Statistics and the nineteenth annual meeting was held at the University of Washington, Seattle, Washington on August 21-24, 1956 in conjunction with the national annual meeting of the Biometric Society, the American Mathematical Society, the Mathematical Association of America, and the Econometric Society. A number of the sessions were joint (and are so designated) with these organizations. All meetings were held on the University of Washington campus. The following 134 members of the Institute attended:

I. J. Abrams, M. S. Ahmed, H. L. Alder, Stephen Allen, C. B. Allendoerfer, A. G. Anderson, F. C. Andrews, K. J. Arnold, R. E. Barlow, C. B. Bell, Z. W. Birnbaum, David Blackwell, J. R. Blum, R. C. Bose, A. H. Bowker, J. V. Breakwell, H. D. Brunk, K. A. Bush, L. D. Calvin, D. G. Chapman, Herman Chernoff, W. S. Connor, D. R. Cox, J. H. Curtiss, Ivonne Cuttle, D. A. Darling, W. J. Dixon, P. J. Doyle, J. A. Dudman, Churchill Eisenhart, Benjamin Epstein, H. P. Evans, T. S. Ferguson, Martin Fox, J. S. Frame, R. S. Gardner, D. W. Gaylor, H. M. Gehman, Dorothy Morrow Gilford, W. A. Golowski, F. A. Graybill,

Geoffrey Gregory, John Gurland, Donald Guthrie, Bernard Harris, T. E. Harris, L. L. Helms, P. G. Hoel, R. V. Hogg, Harold Hotelling, A. S. Householder, M. Iqbal, Walter Jacobs, John L. Jaech, P. W. M. John, H. L. Jones, Samuel Karlin, E. S. Keeping, O. M. Klose, C. F. Kossack, C. H. Kraft, William Kruskal, G. M. Kuznets, R. B. Leipnik, Jerome C. R. Li, G. J. Lieberman, S. P. Lloyd, A. T. Lonseth, F. W. Lott, R. C. McCarty, W. G. Madow, Frank Massey, N. U. Mayall, P. L. Meyer, D. F. Mills, Alex Mood, R. A. Moore, Lincoln Moses, Stanley W. Nash, J. Neyman, G. E. Nicholson, Jr., D. B. Owen, Mohan Pavate, M. P. Peisakoff, R. S. Pinkham, G. B. Price, Ronald Pyke, Howard Raiffa, P. H. Randolph, R. R. Read, G. J. Resnikoff, D. L. Richter, Gerald Rogers, S. N. Roy, Herman Rubin, R. C. Schneider, Lorraine Schwartz, Elizabeth L. Scott, Franklin Sheehan, M. M. Siddiqui, W. L. Smith, G. P. Steck, Charles Stein, Rothwell Stephens, A. D. Stewart, R. F. Tate, W. F. Taylor, Henry Teicher, E. A. Thomas, G. B. Thomas, Jr., F. H. Tingey, F. H. Trinkl, D. R. Truax, J. R. Vatnsdal, Elizabeth Vaughan, R. E. Walpole, J. E. Walsh, L. H. Wegner, J. G. Wendel, Oscar Wesler, Kathleen White, Zivia S. Wurtele, R. K. Zeigler.

The program of the meeting was as follows:

TUESDAY, AUGUST 21, 1956

10:00 a.m. Invited Papers I.

Place: Room 320, Physics Hall

Chairman: A. M. MOOD, General Analysis Corporation

1. *The Asymptotic Attainment of Bayes Risk*, DAVID BLACKWELL, University of California, Berkeley.
2. *Some Problems in Asymptotic Theory*, LUCIEN LE CAM, University of California, Berkeley.

11:30 a.m. Special Invited Address.

Place: Room 320, Physics Hall

Chairman: WILLIAM KRUSKAL, University of California, Berkeley

Asymptotic Theory of Kolmogorov, Smirnov, and von Mises Type Statistics,
DONALD A. DARLING, University of Michigan.

2:00 p.m. American Mathematical Society Colloquium Lecture.

Place: Health Sciences Auditorium

Speaker: SALOMON BOCHNER, Princeton University

Title: *Harmonic Analysis and Probability*.

3:00 p.m. Applications to Physical Sciences.

Place: Room 320, Physics Hall

Chairman: HAROLD HOTELLING, University of North Carolina, Chapel Hill

1. *Problem of Rotation of Galaxies of Different Types—Statistical Aspects*, N. U. MAYALL, Lick Observatory.
2. *Internal Motions in Gaseous Masses of Cosmical Dimensions*, GUIDO MUNCH and O. C. WILSON, Mount Wilson and Mount Palomar Observatories, and California Institute of Technology.
3. *Review of Certain Astronomical Problems and Their Statistical Treatments*, J. NEYMAN and ELIZABETH L. SCOTT, University of California, Berkeley.
4. *Use of the r^{th} Brightest Star in a Galaxy as a Distance Indicator*, MANDAKINI SANE, University of California, Berkeley.
5. *Distribution of the Number of Droplets in Unit Lengths of a Track of a Cosmic*

Ray Particle in a Cloud Chamber, ROBERT READ, University of California, Berkeley.

6. *Effect of Expansion of the Universe on the Serial Correlation of Counts of Images of Galaxies in Regularly Spaced Squares—A Simplified Model*, MARTIN FOX, University of California, Berkeley.

WEDNESDAY, AUGUST 22, 1956

9:00 a.m. Invited Papers II.

Place: Room 320, Physics Hall

Chairman: MRS. BERNICE BROWN, The RAND Corporation

1. *Confidence Regions for Dependent Regression*, PAUL G. HOEL, University of California, Los Angeles.
2. *Functional Relationships with All Variables Subject to Error*, JOHN GURLAND, Iowa State College.

9:00 a.m. Statistical Problems in Medicine and Biology. (Joint with the Biometric Society)

Place: Room 334, Physics Hall

Chairman: W. TAYLOR, School of Aviation Medicine, Randolph Air Force Base

1. *Some Nonparametric Techniques*, L. MOSES, Stanford University, Stanford.
2. *An Investigation of the Log Transformation of Growth Data*, W. BECKER, University of California, Berkeley, and Western Washington Experimental Station, Puyallup.
3. *Reaction Rates in Geometrically Constrained Enzyme Systems*, D. JENDEN, Naval Medical Research Institute, Bethesda, and University of California at Los Angeles.

10:30 a.m. Invited Papers III.

Place: Room 320, Physics Hall

Chairman: BENJAMIN EPSTEIN, Wayne University and Stanford University

1. *Law of Small Numbers*, WILLIAM KRUSKAL, University of California, Berkeley.
2. *Optimal Multivariate Tests*, CHARLES STEIN, Stanford University.
3. *What Judgments are Sufficient for Statistics?* I. J. GOOD, Cheltenham, England.

1:30 p.m. Biometric Society Special Invited Address.

Place: Room 320, Physics Hall

Chairman: J. NEYMAN, University of California, Berkeley

Models and General Mathematical Principles in Biology and Sociology, N. RA SHEVSKY, University of Chicago.

1:45 p.m. Mathematical Problems in Incomplete Block Designs.

Place: Room 334, Physics Hall

Chairman: BURTON W. JONES, University of Colorado

1. *Recent Advances in Partially Balanced Designs*, R. C. BOSE, University of North Carolina, Chapel Hill.
 2. *Symmetrical Balanced Designs*, H. J. RYSER, Ohio State University.
- Discussant: W. J. CONNOR, National Bureau of Standards.

8:00 p.m. Council Meeting.

Place: Room 320, Physics Hall

THURSDAY, AUGUST 23, 1956**9:00 a.m. Prediction Problems. (Joint with Biometric Society)**

Place: Room 320, Physics Hall

Chairman: DAVID BLACKWELL, University of California, Berkeley

1. *New Light on the Multiple Correlation Coefficient*, HAROLD HOTELLING, University of North Carolina, Chapel Hill.
2. *Optimal Estimates of Multiple Criteria with Restrictions on the Covariance Matrix of Estimated Criteria*, PAUL HORST, University of Washington.
3. *Procedural Considerations in Forecasting Populations*, V. A. MILLER, University of Washington.

10:00 a.m. Inventory Policy and Dynamic Programming. (Joint with Econometric Society)

Place: Room 314, Physics Hall

Chairman: GEORGE DANTZIG, The RAND Corporation

1. *A Note on the Optimal Character of the (s, S) Policy in the Inventory Problem*, JACK ABRAMS, University of California, Berkeley.
2. *Optimal Sequential Search Problems*, SELMER JOHNSON, The RAND Corporation.
3. *Ordering Policy for Poisson Determined Supply and Demand*, S. ALLEN and G. FEENEY, Stanford Research Institute.
4. *The Min-Max Solution of a One-stage Inventory Problem (20 minutes)*, HERBERT SCARF, The RAND Corporation.

11:00 a.m. Invited Papers IV.

Place: Room 334, Physics Hall

Chairman: W. J. DIXON, University of California, Los Angeles

1. *The Sequential Item Selection Problem in Classification Studies—The Case of Dichotomous Variables*, HOWARD RAIFFA, Center for Advanced Study in the Behavioral Sciences.
2. *On the Use of Concomitant Variables in the Selection of an Experimental Design*, D. R. COX, University of Cambridge and University of North Carolina, Chapel Hill.

2:00 p.m. Invited Papers V.

Place: Room 320, Physics Hall

Chairman: D. R. COX, University of Cambridge and University of North Carolina, Chapel Hill

1. *Transient Queue Phenomena*, WALTER L. SMITH, University of North Carolina, Chapel Hill.
2. *Some Queueing Statistics*, EDGAR REICH, The RAND Corporation and University of Minnesota.
3. *Some Models of Birth and Death Processes—Linear Growth and Queueing Problems*, SAMUEL KARLIN, Stanford University, and JAMES MCGREGOR, California Institute of Technology.

4:00 p.m. Contributed Papers I.

Place: Room 320, Physics Hall

Chairman: C. H. KRAFT, University of California, Berkeley

1. *Efficient Small Sample Nonparametric Median Tests with Bounded Significance Levels*, JOHN E. WALSH, Lockheed Aircraft Corporation.
2. *Bayes Approach to Control of Fraction Defective*, JOHN V. BREAKWELL, North American Aviation, Inc.
3. *The Distribution of the Extreme Mahalanobis' Distance from Sample Mean (Preliminary Report)*, YVONNE G. M. G. (MRS. P. M.) CUTTLE, University of British Columbia, (introduced by S. W. Nash).
4. *On the Moments of Order Statistics from a Normal Population*, R. C. BOSE and SHANTI S. GUPTA, University of North Carolina, Chapel Hill.
5. *A Comparison of the Power Curves of Some Double Sample Tests*, DONALD B. OWEN, Sandia Corporation.
6. *On Some Nonparametric C-sample Tests*, FRED C. ANDREWS, University of Nebraska.
7. *An Asymptotically Distribution-free Multiple Comparison Method with Application to the Problem of r Rankings of m Objects*, IRENE ROSENTHAL and THOMAS S. FERGUSON, University of California, Berkeley.
8. *Some Asymptotic Results on Wald's Approximate Classification Statistic*, M. IQBAL, University of North Carolina, Chapel Hill.
9. *On the Studentized Largest and Smallest Chi-squared*; K. V. RAMACHANDRAN, University of Baroda, India. (By Title)
10. *On the Distribution of Ranks and of Certain Rank Order Statistics*, PROFESSOR MEYER DWASS, Northwestern University and Stanford University. (By Title)
11. *Contributions to Distribution-free Population Comparisons*, WILLIAM E. PERRAULT and WALDO A. VEZEAU, St. Louis University. (By Title)
12. *Validity of Approximate Normality Values for $\mu \pm k\sigma$ Areas of Practical Type Continuous Populations*, JOHN E. WALSH, Lockheed Aircraft Corporation. (By Title)
13. *Maximum Likelihood Estimation of Restricted Parameters (Preliminary Report)*, H. D. BRUNK, University of Missouri. (By Title)
14. *Confidence Intervals for the Number of Cells in a Multinomial Population with Equal Cell Probabilities*, BERNARD HARRIS, Stanford University and Department of Defense. (By Title)

4:00 p.m. Contributed Papers II. (Joint with American Mathematical Society)

Place: Room 131, Bagley Hall

Chairman: I. J. GOOD, Cheltenham, England

1. *The Quadratic Birth Process*, PETER W. M. JOHN, University of New Mexico.
2. *On a Uniqueness Property not Enjoyed by the Normal Distribution*, GEORGE P. STECK, Sandia Corporation.
3. *Moment Generating Functions of Quadratic Forms in Serially Correlated Normal Variables*, R. B. LEIFNIK, University of Washington.
4. *Solution of a Ranking Problem from Paired Comparisons*, L. R. FORD, JR., The RAND Corporation.
5. *Coincidence Probabilities (Preliminary Report)*, SAMUEL KARLIN, Stanford University, and J. L. MCGREGOR, California Institute of Technology.
6. *A Mean Martingale Convergence Theorem*, L. L. HELMS, Convair.

7. *Stochastic Convergence of Semimartingales*, KLAUS KRICKEBERG, University of Wisconsin.
8. *A General Convergence Theorem for Sequences of Stochastic Processes*, E. G. KIMME, Oregon State College. (By Title)
9. *Almost Sure Everywhere Divergence of Random Series*, ARYEH DVORETSKY, Hebrew University, Jerusalem, and Columbia University.

6:00 p.m. Business Meeting.

Place: Room 320, Physics Hall

8:00 p.m. Council Meeting.

Place: Room 320, Physics Hall

FRIDAY, AUGUST 24, 1956

9:00 a.m. Invited Papers VI.

Place: Room 320, Physics Hall

Chairman: LEO KATZ, Michigan State University

1. *Law of Small Numbers*, WILLIAM KRUSKAL, University of California, Berkeley.
2. *Some Non-parametric Tests for Independence*, JULIUS R. BLUM, Indiana University.
3. *Some Non-parametric Generalizations of Analysis of Variance and Multivariate Analysis*, S. N. ROY, University of North Carolina, Chapel Hill.

9:00 a.m. Stochastic Population Problems. (Joint with Biometric Society)

Place: Room 334, Physics Hall

Chairman: DOUGLAS CHAPMAN, University of Washington, Seattle.

1. *A Stochastic Model for the Tunneling and Retunneling of Flour Beetles*, M. AHMED, University of California, Berkeley.
2. *A Stochastic Model for the Number of Beetles on the Surface of Flour*, EARL R. RICH, University of California, Berkeley.

11:00 a.m. Invited Papers VII.

Place: Room 320, Physics Hall

Chairman: DONALD R. TRUAX, California Institute of Technology

1. *The Distributions of Shadows with Applications to Traffic and Counter Problems*, HERMAN CHERNOFF, Stanford University.
2. *Bounds for Stochastic Processes*, Z. W. BIRNBAUM, University of Washington.
3. *Quasi-Martingales and Stochastic Integrals*, HERMAN RUBIN, University of Oregon.

1:30 p.m. Contributed Papers III.

Place: Room 320, Physics Hall

Chairman: M. R. MICKEY, The RAND Corporation

1. *Incomplete Sufficient Statistics and Similar Tests*, ROBERT A. WIJSMAN, University of California, Berkeley, (introduced by David Blackwell).
2. *Multi-decision Problems for the Multivariate Exponential Family*, DONALD R. TRUAX, California Institute of Technology.
3. *Some Distributions Related to $D_n +$* , Z. W. BIRNBAUM and R. PYKE, University of Washington.

4. *Sequential Distribution-free Tolerance Regions*, SAM C. SAUNDERS, University of Washington.
5. *Idempotent Matrices and Quadratic Forms in the General Linear Hypothesis*, FRANKLIN A. GRAYBILL and GEORGE MARSAGLIA, Oklahoma A. and M. College.
6. *On Infinitely Divisible Random Vectors*, MEYER DWASS, Northwestern and Stanford Universities and HENRY TEICHER, Purdue and Stanford Universities.
7. *Further Contributions to Multivariate Confidence Bounds*, S. N. ROY and R. GNANADESIKAN, University of North Carolina, Chapel Hill.
8. *Contributions to univariate and multivariate components of variance analysis*, S. N. ROY and R. GNANADESIKAN, University of North Carolina, Chapel Hill.
9. *The Linear Hypothesis, Information, and the Analysis of Variance*, (Preliminary Report), CHESTER H. MCCALL, JR., The George Washington University. (By Title)
10. *A Sequential Multiple Decision Procedure for Selecting the Multinomial Event with the Largest Probability (Preliminary Report)* R. E. BECHHOFFER, Cornell University, and M. SOBEL, Bell Telephone Laboratories. (By Title)
11. *On the Existence of Uniformly Efficient Estimates*, R. R. BAHADUR, University of Chicago. (By Title)
12. *Definite Quadratic Forms and Discontinuous Factor*, ANDRÉ G. LAURENT, Michigan State University. (By Title)
13. *A Further Contribution to the Theory of Univariate Sampling on Successive Occasions (Preliminary Report)*, B. D. TIKKIWAL, University of North Carolina and Karnatak University. (By Title)
14. *Invariance, Sequential Decision Functions, and Continuous Time Processes*, J. KIEFER, Cornell University. (By Title)
15. *On the Construction of Fractional Factorial Designs*, ROBERT C. BURTON, National Bureau of Standards, (introduced by W. C. CONNOR). (By Title)

CHARLES KRAFT
Associate Secretary

MINUTES OF THE ANNUAL BUSINESS MEETING 1956

A business meeting of the Institute of Mathematical Statistics was called to order at 6:00 p.m., August 22, 1956, in Room 320, Physics building, University of Washington, Seattle, by President David Blackwell. Approximately 42 persons were present. A special announcement was made by G. E. Nicholson about the tragic death of John Wishart.

Minutes of the Annual Business Meeting held in New York in December, 1955, were read and approved. Z. W. Birnbaum introduced a resolution which had been recommended by the Council that the Kingston Policy on holding unsegregated meetings be made the permanent policy of the Institute. After considerable discussion this motion was passed.

The reports of the Editor, Treasurer, Secretary, and Program coordinator were presented and approved. It was announced by the Secretary that Dorothy

Morrow Gilford had been appointed by the Council to be Associate Secretary for the Eastern Region to fill Allan Birnbaum's unexpired term.

The tellers were instructed to accept ballots from members who had not relayed them by mail.

The President presented his report and turned the chair over to new President Alexander Mood. President Mood extended the thanks of the Institute to the outgoing president for his work for the Institute during the past year.

Lucien LeCam moved that meetings of the Council be open for all members of the Institute to observe and listen. This motion passed.

W. J. Dixon introduced a resolution of thanks to the University of Washington administration and the arrangements committee for the meeting. This was passed.

The tellers announced the election of the following:

President-Elect L. J. Savage

Members of I.M.S. Council for term 1956-1959:

T. W. Anderson

M. S. Bartlett

J. Berkson

Erich Lehmann

The meeting was adjourned at 8:15 p.m.

GEORGE E. NICHOLSON, JR.
Secretary

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1956

The affairs of the Institute ran smoothly during 1956.

We have 1649 members as compared with 1505 a year ago. A substantial part of the increase can be attributed to the work of this year's Committee on Individual Memberships, under the chairmanship of Benjamin Epstein. The Council voted special thanks to this committee for an outstanding job.

We expect the *Annals* to reach a record size of 1225 pages this year, a considerable increase over last year. In spite of this increase, we anticipate that the Institute will, financially, about break even.

The Council has voted to hold a Summer Institute in 1957 on analysis of variance provided we get a grant from the National Science Foundation for this purpose. The organizing committee for the 1957 summer institute, with T. W. Anderson as chairman, is continuing its work.

In addition to the annual meeting at Seattle on August 21-24, with the American Mathematical Society, the Institute held an Eastern Regional meeting in Princeton on April 20-21 and a Central Regional meeting at the University of Chicago on April 27-28. Future meetings already being contemplated are a Central Regional meeting in Detroit on September 7-9, a special meeting in conjunction with the AAAS in New York during the Christmas 1956 holidays,

an Eastern Regional meeting at Catholic University on March 7-9, 1957, the annual 1957 meeting with the American Statistical Association in Atlantic City early in September, 1957, and the annual 1958 meeting with the American Mathematical Society, the time and place of which have not been determined.

Our committee on Professional Standards, under the chairmanship of B. F. Kimball, has prepared a letter, which it plans to send to various personnel officers in state and local governments, concerning appropriate standards for statisticians in government work.

We were happy to be able to invite five distinguished Russian probabilists, Kolmogorov, Hincin, Linnik, Gnedenko, and Prohorov, to attend our Seattle meeting. Unfortunately, because of delays in official machinery, the invitations could not be transmitted until July 2, and none of the Russians was able to come on such short notice. Kolmogorov wrote a most cordial letter, heartily endorsing the desire expressed in our letter to establish a closer contact between scholars of our two countries, and expressing the hope that some Soviet probabilists would be able to attend a later meeting of the Institute of Radio Engineers on information theory at MIT on September 10-12.

It is a pleasure to announce that the Rietz Lecture Committee, with J. Neyman as chairman, has designated J. Wolfowitz as our Rietz lecturer for 1957. He will deliver the lecture at the 1957 annual meeting in Atlantic City.

The Council has unanimously recommended, and the members present at the 1956 membership meeting have unanimously voted, that the Kingston policy be the permanent policy of the Institute. This policy is: "It is the policy of the Institute of Mathematical Statistics that all its meetings shall be held on a completely non-segregated basis. In particular prior to determining the place of a forthcoming meeting, the Secretary of the IMS shall ascertain that meeting halls, eating facilities and housing accommodations adequate for the expected attendance will be available on a non-segregated basis, and that all social events connected with the meetings shall be non-segregated." We have not experienced, and do not anticipate, substantial difficulty in conforming to this policy.

We have reached an agreement with the University of Chicago Press, under which the Institute will cooperate with the Press in publishing, at no cost to the Institute, a series of statistical monographs, which will be available to IMS members at a $\frac{1}{3}$ discount on publication orders. You will be notified when monographs are to appear.

You have already been notified of the arrangement with the University of California Press, under which IMS members may purchase volumes of the *Proceedings of the Third Berkeley Symposium on Probability and Statistics* at a 25 per cent discount.

The Committee on Special Invited Papers, under the chairmanship of William Kruskal, has arranged for the following three Special Invited Papers:

1. W. J. Youden, "Experimental Designs for Industrial Research"; scheduled for Central Regional Meeting, Detroit, September, 1956.
2. D. L. Wallace, on asymptotic approximations to distributions; scheduled for Annual Meeting, Atlantic City, September, 1957.

3. C. Hildreth, on statistical problems in economics; as yet unscheduled.

The 1956 nominating committee consists of Charles Stein, chairman, Joseph Berkson, Kai Lai Chung, John Curtiss, and David Kendall.

L. J. Savage was elected president-elect for 1957, and T. W. Anderson, M. S. Bartlett, J. Berkson, E. L. Lehmann were elected to the Council for 1957-59.

A list of 1956 Institute committees and representatives is included as an appendix to this report. I know that all Institute members join me in expressing gratitude to these members for so generously giving their time to the affairs of the Institute.

In closing, I wish to congratulate our nine newly elected Fellows. They are

Allan Birnbaum
D. R. Cox
N. L. Johnson
Samuel Karlin
W. H. Kruskal
Walter L. Smith
Milton Sobel
Erling Sverdrup
E. J. Williams

DAVID BLACKWELL,
President

Appendix. 1956 IMS Committees and Representatives

(The first name is that of the chairman)

Academic Institutional Members: G. J. Lieberman, H. Robbins.

Activities and Development: T. W. Anderson, J. Berkson, A. Bowker, T. E. Harris, W. Kruskal, S. Wilks.

Exchanges: P. Dwyer, T. E. Harris (*ex officio*), G. E. Nicholson, Jr., (*ex officio*).

Fellows: H. Levene, F. Anscombe, M. Bartlett, D. Blackwell, L. Goodman, L. J. Savage.

Finance: M. Spiegelman, A. Bowker, K. J. Arnold.

Individual Membership: B. Epstein, E. Crow, D. Chapman, F. Grubbs, E. Pike.

Non-Academic Institutional Members: Bernice Brown, A. Householder, P. Olmstead, C. C. Hand.

Physical Facilities: Z. W. Birnbaum, Leo Katz, G. E. Nicholson, Jr.

Professional Standards: B. F. Kimball, R. Burgess, C. Eisenhart, G. Harrington, A. Householder, J. Lev, H. Marshall, R. Patton, J. E. Walsh.

Program Committees:

Annual: M. Peisakoff, H. Raiffa, H. Levene, M. Rosenblatt, D. G. Chapman, J. Kiefer, B. Harshbarger, L. Katz (*ex officio*).

Eastern: M. Zelen, R. L. Anderson, R. Bradley, G. Burrows, C. Derman, M. Halperin, L. Weiss, Leo Katz (*ex officio*).

Central: S. Isaacson, F. C. Andrews, L. Cote, L. Goodman, J. Gurland, I. Olkin.

Western: L. Moses, H. Rubin, Z. Birnbaum, B. Brown, M. Sandomire, G. Lieberman, J. Yerushalmy.

Rietz Lecture: J. Neyman, H. Hotelling, W. Feller.

Special Invited Papers: W. Kruskal, R. L. Anderson, K. Arrow, G. Cox, A. T. Craig, H. Hartley, I. Savage, T. E. Harris (*ex officio*).

Mathematical Tables: J. W. Tukey, R. L. Anderson, A. Bowker, E. Cureton, W. Dixon,

C. Dunnett, C. Eisenhart, J. A. Greenwood, H. Hartley, E. Kaplan, W. Kruskal, D. B. Owen, D. Teichroew, Max Woodbury.
Fast Machines: R. L. Anderson, F. S. Acton, K. J. Arnold, A. S. Householder, C. F. Cosack, W. H. Kruskal, W. J. Merrill, H. A. Meyer, J. Moshman, H. W. Norton, G. J. Resnikoff, R. Slimak, Z. Szatrowski, D. Teichroew.
1957 Summer Institute: T. W. Anderson, H. Scheffé, J. W. Tukey, J. Cornfield, O. Kempthorne.
Inviting Russian Probabilists: E. Lukacs, D. Blackwell, J. L. Doob, J. Neyman, H. Robbins.
1956 Nominating Committee: C. Stein, J. Berkson, K. L. Chung, J. Curtiss, D. G. Kendall.

Representatives

American Association for the Advancement of Science: H. Hotelling (through 1957).
 Intersociety Committee on Standardization of Nomenclature and Symbols: H. Raiffa.
 National Research Council: S. S. Wilks (through 1957).
 Policy Committee for Mathematics: J. F. Daly.
 Council of Population and Housing Census Users: Paul Meier.

REPORT OF THE SECRETARY FOR 1956

During 1956 the Institute held its 69th through 72nd meetings. A business meeting was held during the 71st (19th Annual) meeting. The Program Committees are to be congratulated on the excellent programs which have been arranged under the immediate direction of M. P. Peisakoff, Stanley Isaacson, and Gottfried E. Noether with the overall guidance of our Program Coordinator, Leo Katz. The Assistant Secretaries, M. R. Wilk, D. L. Wallace, D. G. Chaoman, and J. E. Lanahan are to be congratulated on the physical arrangements, and the Associate Secretaries, Allan Birnbaum, William Kruskal, and C. H. Kraft, on their performance of the duties of the Secretary with respect to the meetings.

GEORGE E. NICHOLSON, JR.
Secretary

INTERIM EDITOR'S REPORT FOR 1956

The volume of material submitted to the *Annals* (in terms of manuscript pages) has shown an upward trend. In the year August 1, 1954–July 31, 1955 it was higher than in any preceding year. In the following year, the one just ending, it was somewhat less, but still higher than any of the previous years. The 1956 *Annals* is accordingly being expanded, the Council having authorized 1225 pages for the year; the increase in size will enable the backlog of accepted papers to be brought below one issue. The Council has also authorized a 1957 volume of 1100 pages.

Thanks are due to many people, other than the regular members of the editorial staff, for generous refereeing assistance and other help. Specific acknowledgment will be made in the final report for the 1956 volume, which will appear in the March, 1957 issue.

PUBLICATIONS RECEIVED

- HERDAN, G., *Language as Choice and Chance*, P. Noordhoff, Ltd., Publishers, Groningen, Holland, \$8.00.
- LANING, J. HALCOMBE, JR., and RICHARD H. BATTIN, *Random Processes in Automatic Control*, McGraw-Hill Series in Control Systems Engineering, McGraw-Hill Book Company, Inc., New York, 1956, 434 pp., \$10.00.
- PETTERSEN, SVERRE, *Weather Analysis and Forecasting*, Second Edition, Volume I, Motion and Motion Systems, McGraw-Hill Book Company, Inc., New York, 1956, 428 pp. \$8.50.
- The Biological Effects of Atomic Radiation, A Report to the Public*, National Academy of Sciences—National Research Council, Washington, D. C., 1956, 40 pp.
- The Biological Effects of Atomic Radiation, Summary Reports*, National Academy of Sciences—National Research Council, Washington, D. C., 1956, 108 pp.

INSTITUTIONAL MEMBERS

- BELL TELEPHONE LABORATORIES, INC., TECHNICAL LIBRARY, 463 West Street, New York 14, New York.
- INTERNATIONAL BUSINESS MACHINES CORPORATION, New York
- IOWA STATE COLLEGE, STATISTICAL LABORATORY, Ames, Iowa
- MASSACHUSETTS INSTITUTE OF TECHNOLOGY, HAYDEN LIBRARY, PERIODICAL DEPARTMENT, Cambridge 39, Massachusetts
- MICHIGAN STATE COLLEGE, DEPARTMENT OF MATHEMATICS, East Lansing, Michigan
- PRINCETON UNIVERSITY, DEPARTMENT OF MATHEMATICS, SECTION OF MATHEMATICAL STATISTICS, Princeton, New Jersey
- PURDUE UNIVERSITY LIBRARIES, Lafayette, Indiana
- STATE UNIVERSITY OF IOWA, Iowa City, Iowa
- UNIVERSITY OF CALIFORNIA, STATISTICAL LABORATORY, Berkeley, California
- UNIVERSITY OF ILLINOIS, Urbana, Illinois
- UNIVERSITY OF NORTH CAROLINA, INSTITUTE OF STATISTICS, Chapel Hill, North Carolina
- UNIVERSITY OF WASHINGTON, LABORATORY OF STATISTICAL RESEARCH, Seattle, Washington

ESTADISTICA

Journal of the Inter American Statistical Institute

Volume XIV, No. 52

Contents

September 1956

ARTICLES

- Aplicaciones Estadísticas en las Ciencias Físicas en los Estados Unidos (traducción)
WILLIAM R. PABST, JR.
Algunos Errores en la Declaración de Edad en los Censos de Población de 1950 en Centro
América y México..... JORGE ARIAS B.
Aproximación de la Tasa Anual Promedio de Cambio..... JACOB S. SIEGEL
Sources, Procedures of Compilation, and Types of Current Industrial Statistics in Canada
DOMINION BUREAU OF STATISTICS
Programación del Desarrollo Estadístico Nacional
OFICINA DE ESTADÍSTICA DE LAS NACIONES UNIDAS
Collection of Mental Disease Statistics in the United States..... MORTON KRAMER
Problemas en la Aplicación del Muestreo en Encuestas Agropecuarias en la América Latina
OFICINA DE ESTADÍSTICA DE LA FAO
Problemas Encontrados en Estudios de Gastos de la Familia Hechos Recientemente en Países
Latinoamericanos..... PAULINE B. PARO
Special Features. Legal Provisions. International Resolutions relating to Statistics. Institute
Affairs. Statistical News. Publications.

Published quarterly

Annual subscription price \$3.00 (U. S.)

INTER AMERICAN STATISTICAL INSTITUTE

Pan American Union
Washington 6, D.C.

ECONOMETRICA

Journal of the Econometric Society
Contents of Vol. 24, No. 4 - October, 1956

- HOLLIS B. CHENERY AND KENNETH S. KRETSCHMER..... Resource Allocation for Economic Development
JOHN C. H. FEI..... A Fundamental Theorem for the Aggregation Problem of Input-Output Analysis
JEROME K. PERCUS AND LEON QUINTO..... Application of Linear Programming to Competitive Bond
Bidding
P. J. VERDOORN..... Complementarity and Long-Range Projections
HARVEY WAGNER..... An Eclectic Approach to the Pure Theory of Consumer Behavior
CONRAD A. BLYTH..... The Theory of Capital and Its Time Measures
J. D. SARGAN..... A Note on Mr. Blyth's Article
MAURICE McMANUS..... On Hatanaka's Note on Consolidation
E. BURGER..... On the Stability of Certain Economic Systems
BOOK REVIEWS
The Alphabet of Economic Science (Philip H. Wicksteed). Review by George J. Stigler
Elementi di politica economica razionale (Eraldo Fossati). Review by Gerhard Tintner
Business Concentration and Price Policy (NBER). Review by Merton J. Peck
Einführung in die Betriebswirtschaftslehre (M. Lohmann). Review by Eberhard Fels
Theory of Games as a Tool for the Moral Philosopher (R. B. Braithwaite). Review by David Hawkins
Etude Econométrique de la Demande de Talco (Roger-Paul Congard). Review by R. L. Basmann
Monopoly and Competition and Their Regulation (E. H. Chamberlin, Ed.). Review by F. H. Hahn
Die Landwirtschaft in der volkswirtschaftlichen Entwicklung. Eine Betrachtung über Beschäftigung und Ein-
kommen (H. Weber). Review by Kaethe Mengelberg
Marginal-Cost Price-Output (Burnham P. Beckwith). Review by Nancy D. Ruggles
Die Nachfrage nach Nahrungsmitteln und ihre Abhängigkeit von Preis- und Einkommensänderungen: Ein
ökonomische Untersuchung von Wirtschaftsrechnungen Hamburger Angestellten- und Arbeiterhaushaltungen
(Heins Gollnick). Review by Eberhard Fels

Published Quarterly

Subscription rates available on request

The Econometric Society is an international society for the advancement of economic theory in its
relation to statistics and mathematics.

Subscriptions to *Econometrica* and inquiries about the work of the Society and the procedure in applying
for membership should be addressed to Richard Ruggles, Secretary, The Econometric Society, Box
1284, Yale University, New Haven, Connecticut.

BIOMETRIKA

Volume 43

Contents Parts 3 and 4, December 1956

ROYSTON, Erica. Studies in the history of probability and statistics. III. A note on the history of the graphical presentation of data. WILLIAMS, C. B. Studies in the history of probability and statistics. IV. A note on an early statistical study of literary style. WALKER, A. M. A goodness of fit test for spectral distribution functions of stationary time series with normal residuals. GANI, J. Sufficiency conditions in regular Markov chains and certain random walks. DERMAN, C. Some asymptotic distribution theory for Markov chains with a denumerable number of states. LAWLEY, D. N. A general method for approximating to the distribution of likelihood ratio criteria. JAMES, G. S. On the accuracy of weighted means and ratios. BAILEY, N. T. J. On estimating the latent and infectious periods of measles. II. Families with three or more susceptibles. BAILEY, N. T. J. Significance tests for a variable chance of infection in chain-binomial theory. WHITTLE, P. On the variation of yield variance with plot size. WATSON, G. S. and WILLIAMS, E. J. On the construction of significance tests on the circle and the sphere. QUENOUILLE, M. H. Notes on bias in estimation. ROY, S. N. & MITRA, S. K. An introduction to some non-parametric generalisations of analysis of variance and multivariate analysis. KAMAT, A. R. A two-sample distribution-free test. BARTON, D. E. Addendum. The limiting distribution of Kamat's test statistics. RAY, W. D. Sequential analysis applied to certain experimental designs in the analysis of variance. BROADBENT, S. R. Lognormal approximation to products and quotients. BLISS, C. I., COCHRAN, W. G. and TUKEY, J. W. A rejection criterion based upon the range. CROW, E. L. Confidence intervals for a proportion.

Miscellanea

Contributions by ANSCOMBE, F. J., BLANK, A. A., COX, D. R., DAVID, F. N., DAVID, H. A., DOUGLAS J. B., GILBERT, N. E. G., HUTTLY, N. A., MOORE, P. G., REIERSØL, O., RUBEN, H., STUART, A., TAYLOR, J., WATSON, G. S.

Corrigenda

GANI, J.; and *Tracts for Computers XXVI*

Reviews

Other Books Received

The subscription price, payable in advance, is £5s. inland, 54s. export (per volume including postage). Cheques should be drawn to Biometrika and sent to "The Secretary, Biometrika Office, Department of Statistics, University College, London, W.C. 1." All foreign cheques must be in sterling and drawn on a bank having a London agency.

MATHEMATICAL REVIEWS

A journal containing reviews of the mathematical literature of the world, with full subject and author indices

Publication of this journal is sponsored by the American Mathematical Society, Mathematical Association of America, Institute of Mathematical Statistics, London Mathematical Society, Edinburgh Mathematical Society, Union Matematica Argentina, and others.

Subscriptions accepted to cover the calendar year only.
Issues appear monthly except July. \$20.00 per year.

Send subscription order or request for sample copy to

AMERICAN MATHEMATICAL SOCIETY
80 Waterman Street, Providence 6, Rhode Island

JOURNAL OF THE ROYAL STATISTICAL SOCIETY

Series B (Methodological)

Vol. XVIII, No. 1, 1956

Some Tests of Significance with Ordered Variables	F. N. DAVID AND N. L. JOHNSON. (With Discussion).
Economic Choice of the Amount of Experimentation	P. M. GRUNDY, M. J. R. HEALY, AND D. H. REES. (With Discussion).
On a Test of Significance in Pearson's Biometrika Table (No. 11)	R. A. FISHER.
A Test of Significance for an Unidentifiable Relation	P. A. P. MORAN.
Confidence Limits for the Gradient in the Linear Functional Relationship	MONICA A. CREASY.
Algebraic Theory of the Computing Routine for Tests of Significance on the Dimensionality of Normal Multivariate Systems	F. E. BINET AND G. S. WATSON.
Some Notes on Ordered Random Intervals	D. E. BARTON AND F. N. DAVID.
A Sequential Test for Randomness	D. J. BARTHOLOMEW.
Partitions in More Than One Dimension	J. H. BENNETT.
On the Estimation of Small Frequencies in Contingency Tables	I. J. GOOD.
An Elementary Method of Solution of the Queueing Problem with a Single Server and Constant Parameters	R. R. P. JACKSON. D. G. CHAMPERNOWNE.
Random Queueing Processes with Phase-type Service	R. R. P. JACKSON.
The Within-Animal Bioassay with Quantal Responses	P. J. CLARINGBOLD.

The Royal Statistical Society, 21, Bentinck Street, London, W. 1

SKANDINAVISK AKTUARIETIDSKRIFT

1955 - Parts 3 - 4

Contents

B. W. CONOLLY	Unbiased Premiums for Stop-Loss Reinsurance
SVEN RIDDERSTRÖM	On Ratio Estimates in Simple Random Sampling with Some Practical Applications
ERICH MICHALUP	Some Approximation Formulae of the Effective Rate and the Force of Interest
IVO LAH	Das Restglied der Taylorschen Reihe des Rentenbarwertes und einige Formeln des Zinsfussproblems für grosse Zinsspannungen
S. VAJDA	Analytical Studies in Stop-Loss Reinsurance. II
G. P. M. HESELDEN	Some Inequalities Satisfied by Incomplete Beta Functions
CARL PHILIPSON	A Tentative Application of the Collective Risk Theory to Crop Insurance
Eksamen i Forsikringsvidenskab og Statistik ved Københavns Universitet	
Översikt av utländska aktuarietidskrifter	

Annual subscription: \$5.00 per year

Inquiries and orders may be addressed to the Editor,

GRANHÄLLSVÄGEN 35, STOCKSUND, SWEDEN

SANKHYĀ

The Indian Journal of Statistics
Edited by P. C. Mahalanobis

Vol. 17, Part 2, 1956

On The Recovery of Inter Block Information Varietal Trials.....	C. RADHAKRISHNA RAO
Classification And Analysis of Linked Block Designs.....	J. ROY AND R. G. LAHA
On the Dual of a PBIB Design and a New Class of Designs with Two Replications.....	C. S. RAMAKRISHNAN
Fractional Replication in Asymmetrical Factorial Designs and Partially Balanced Designs.....	I. M. CHAKRAVARTI
A General Class of Quasifactorial and Related Designs.....	C. RADHAKRISHNA RAO
Two Associate Partially Balanced Designs Involving Three Replications.....	J. ROY AND R. G. LAHA
Some Series of Balanced Incomplete Block Designs.....	D. A. SPROTT
The Concept of Asymptotic Efficiency.....	D. BASU
A Note on the Determination of Optimum Probabilities in Sampling without Replacement.....	DES RAJ

ANNUAL SUBSCRIPTION: 30 rupees (\$10.00), 10 rupees (\$3.50) per issue.
BACK NUMBERS: 45 rupees (\$15.00) per volume; 12/8 rupees (\$4.50) per issue.
Subscriptions and orders for back numbers should be sent to

STATISTICAL PUBLISHING SOCIETY
204/1 Barrackpore Trunk Road
Calcutta 35, India

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized December 12, 1906)

MEMBERS

President: M. Mood, General Electric Corporation, 710 Wilshire Blvd.,
Beverly Hills, California

Vice-President: J. Savage, Committee on Statistics, University of Chicago, Chicago 37,
Illinois

Secretary: W. H. Mendenhall, Department of Statistics, University of North Carolina,
Chapel Hill, North Carolina

Member: H. Bowker, Department of Statistics, Stanford University, Stanford,
California

Member: T. E. Harris, The Rand Corporation, 450 Main Street, Santa Monica,
California

The purpose of the Institute of Mathematical Statistics is to encourage the
development, dissemination, and application of mathematical statistics.

Membership dues including a subscription to the ANNALS OF MATHE-
MATICAL STATISTICS are \$10.00 per year for members of the United States
and Canada and \$15.00 per year for members of other countries. There are special
rates for students and for some other classes of members. Questions regarding
membership in the Institute should be sent to the Secretary of the Institute.

STANDARD FORMS OF WRITING

Standard forms of writing. The difference between capital and small letters should be clearly
marked; care should be taken to avoid confusion between the letter O and the letter 0,
the numeral 1 and the letter l, capital I and the letter i, and other (C) capital and a,
kappa and k, mu and u, nu and v, eta and e, theta and t, etc. Capital letters should be
clearly below or above the line, then above or below (italics) and underlined
letters (e.g., x) are difficult to print and some of them (e.g., x) are frequently
misread by the printer with small letters. Capital letters and small letters may
be indicated by underlining or by other means. Capital letters, boldface letters and
underlined letters are not recommended. Capital letters should be represented with the
usual exp. In writing capital letters, the use of the radical sign is preferable to the radical
sign. Fractions in the body of a sentence should be written as fractions and fractions
referring to the numerator or denominator of a fraction are usually written with the
numerator and denominator in parentheses, e.g., $(a+b)/(c+d)$ rather than $a+b/c+d$.

Authors will ordinarily receive only one proof of their manuscript without charge. A second proof will be
furnished free. Additional copies of the proof will be furnished at cost.

